

THIS WEEK

EDITORIALS

COSMOLOGY Computer simulation opens window on evolution of the Universe **p.134**

WORLD VIEW The regulatory threat to the promise of synthetic biology **p.135**

HIGH FASHION NASA announces winning spacesuit design **p.139**



Tribal gathering

Before they can construct a cell, researchers in synthetic biology must first build bridges between disciplines.

Although ‘-ology’ denotes a subject of study, synthetic biology is more about building than observation. The field’s founding papers were published just 14 years ago, but the scientists who call themselves synthetic biologists already have disparate goals. By many accounts, ‘synbio’ is less a coherent discipline than a collection of tribes under the same name. They do not interact with each other nearly enough.

Some synthetic biologists design genetic circuits to make microbes do fun or useful tricks — blink in sync, count pulses of chemicals, pump out biofuels or detect environmental pollutants. Others revise the chemistry of life’s construction materials, crafting DNA and proteins that contain non-natural building blocks. Still others stitch together stretches of DNA to create whole chromosomes. In 2010, scientists replaced a bacterium’s genome with a synthetic replica. In 2014, a team fabricated a next-generation yeast chromosome, dropping cumbersome sequences and inserting convenient ones.

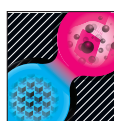
And those who influence the field with calls for funding and regulation are also driven by different goals. They agree that the outcomes of synthetic biology cannot yet be predicted, but some put caution foremost, seeking to avert any disasters that might result from an engineered microbe, whereas others focus on rewards, imagining revolutions in energy, medicine, nutrition or manufacturing.

So far, the field has avoided many of the public-image pitfalls that beleaguered genetically modified organisms (GMOs). Today, the term GMO is identified more with patented crops that help agriculture companies to sell products than with potential tools to grow food more efficiently or generate life-saving drugs. The best known application of synthetic biology, by contrast, is probably the manufacture of an accessible life-saving malaria medicine.

Although synthetic biologists need to address public concerns, they must also tend to the field from within, by fostering connections across disciplines. Divisions run deep. A synthetic biologist may come to a project as an iterative tinkerer, a methodical engineer or an intuitive explorer. Engineering types can be flummoxed by unpredictability, and sciencey types frustrated by demands for inflexible definitions and standards.

Some researchers have focused on establishing standards so that various ‘gear parts’ can be easily interchanged and combined. Others resist, arguing that variability among cells and circuits defies human-devised specifications. Some scientists push for open access, the better to mix and match inventions; others call for strong protections for intellectual property, the better to incentivize development of useful applications. The paucity of connection between these groups has provided ample fodder for sociologists of science.

As part of this week’s special issue, a News Feature on page 152 suggests a way forward: open access and the safe-guarding of intellectual property could work together in a ‘diverse ecology’ in which circuit



BEYOND DIVISIONS

The future of synthetic biology
nature.com/synbio

parts are freely available and built-up components are eligible for patenting. Other gaps could be filled by technological and cultural innovations. Streamlined techniques can now characterize components across many contexts,

including different cell types and combinations. These are helping to bring some predictability to the bewildering variability that is inherent in a living cell. Tellingly, one of the best known synthetic-biology companies, Amyris in Emeryville, California, combines randomized and rational approaches. It screens both randomly generated yeast mutants and designed strains every week to incrementally improve yields of a chemical product.

Synthetic biologists suggest that the various camps fail to connect because of apathy and habit rather than outright hostility. Whatever the reason, the discipline suffers from its divisions. Asked how to overcome the obstacles facing synthetic biology, experts call

“There will probably always be many tribes in synthetic biology, but there is reason to believe the field can avoid tribalism.”

for tighter integration between disciplines and cultures (page 155). Means to this end might include the annual International Genetically Engineered Machine (iGEM) competition or ‘build-a-genome’ courses, which produced the first complete synthetic yeast chromosome earlier this year. Although these programmes arose from an open-access tribe within synthetic biology, they reach undergraduate and even high-school students — nascent scientists not yet

ensconced in a specialty. Perhaps they will bring new and useful kinds of thinking into the field.

Establishment of an international synthetic-biology society might be another way to foster productive mingling while allowing the field to maintain itself as a coherent discipline. Some encouraging developments have been evident at recent conferences, with panels comprising members from various camps. And more research is needed to learn how, for instance, intellectual-property and open-access schemes can be combined to boost innovation.

Although there will probably always be many tribes in synthetic biology, there is reason to believe that the field can avoid tribalism. Indeed, it is now more vital than ever that synthetic biologists present a united front. As Volker ter Meulen, co-chairman of IAP — the Global Network of Science Academies — writes in a World View on page 135, storm clouds are gathering on the horizon. Not everyone agrees that synthetic biology is a force for good, and that opposition has found its voice in a consultation for the global Convention for Biological Diversity. It is crucial, says ter Meulen, that the balancing voice of science is heard before false assumptions lead to the creation of onerous and unnecessary regulation. Everyone can agree on that. ■

DAVID AUSSERHOFER/LEOPOLDINA



Time to settle the synthetic controversy

If synthetic biology is to thrive, the world needs to decide now how the field should be regulated and supported, says **Volker ter Meulen**.

The creation of an artificial yeast chromosome shows that synthetic biology is getting closer to what most scientists want: to be able to deliver benefits to society. The field has already found cheaper ways to produce medicines, and is making progress in applications from water purification to materials design.

The topic is, however, controversial, and that is jeopardizing its promise. Environmental groups argue that it poses risks to health and the environment and have called for a global moratorium. We have been here before: exaggerated fears and uncritical acceptance of claims of the risks of genetic modification led to excessively cautious regulation and a block on innovation that not only slowed the development of new products, but also deterred basic science.

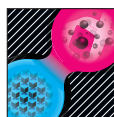
The debate over synthetic biology is now entering a critical phase. The Conference of the Parties to the Convention on Biological Diversity (CBD) — the global framework that governs the protection of biodiversity — is currently exploring possible restrictions and will clarify its position at meetings next month and in October. But given the precedent of how the issue of genetically modified crops was handled, many scientists are worried that some policy-makers will take unsubstantiated concerns of environmental groups at face value and impose cumbersome and unnecessary rules. To prevent that, we need an objective, evidence-based and balanced assessment of the risks and benefits, both within and beyond the CBD. And that means that the voice of science must be heard.

To kick-start the necessary dialogue, the IAP — the Global Network of Science Academies — has published a position statement (see go.nature.com/tmvhf8), drawn from the work of member institutes across Europe and the United States. We (I am co-chair) hope that the statement will spur debate on how best to feed scientific evidence that emerges from peer-reviewed research into the development of policy. This includes the regulations for overseeing research and innovation, as well as, for instance, investment in relevant infrastructure and training.

The IAP represents 106 academies worldwide and wants to have a more active role in global policy issues; last year, for instance, we published a similar position statement on antimicrobial-drug resistance.

In the case of synthetic biology, the world needs to commit to addressing several priorities. First, the scope of synthetic biology needs to be determined. We describe it as the construction of customized biological systems to perform new and improved functions, through the application of principles from engineering and chemical synthesis.

The goal might be new, but many of the techniques are borrowed



BEYOND DIVISIONS

The future of synthetic biology
nature.com/synbio

from existing fields, such as genetic modification. This means that the work is not proceeding entirely without regulation, as some claim; much of it is, in fact, governed by existing rules.

The use, release and movement across borders of genetically modified organisms, for instance, are covered by the application of the Cartagena Protocol on Biosafety.

The recognition that key methods are already controlled is crucial, because it should defuse some of the public controversy about risk. Also important is striking the right balance between statutory regulation and self-governance by scientists and scientific bodies. (The IAP and others have published recommendations on how to develop individual and institutional codes of conduct.)

A second aspect that must be considered more broadly is how the results of synthetic biology are owned and shared. The current situation reflects its mixed parentage from both the biosciences (with its tradition of patenting) and engineering and software development (which embrace open sourcing and sharing). The announcement of how researchers worldwide worked to produce a synthetic yeast chromosome shows how openness can pay off in academia.

As synthetic biology progresses, techniques and tools will inevitably be developed that are not covered by existing regulations. It is reasonable to assume, the IAP argues, that these will allow the research to be done with greater precision. More-controlled modifications to genetic sequences, cells and organisms will facilitate characterization and bring the prospect of reducing unex-

pected and unwanted side effects. Future synthetic-biology techniques and products should therefore be easier to regulate, manage and audit than earlier, less controlled genetic-modification techniques.

Finally, the IAP says that funding bodies across the world must anticipate the potential of synthetic biology and invest in the research, and in the researchers involved. The investment should also incorporate projects in the social sciences and the humanities, which can, for instance, look at concerns about biologists 'creating life' and find better ways to communicate the issues.

Together, these steps should help to ensure that policies on synthetic biology set out sensible practices to mitigate the risk that is inherent in any major advance, yet are flexible enough to encourage research and innovation. ■

Volker ter Meulen is co-chair of the IAP — the Global Network of Science Academies in Trieste, Italy, and past president of the German Academy of Sciences Leopoldina.
e-mail: volker.termeulen@mail.uni-wuerzburg.de

**FUNDING
BODIES ACROSS THE
WORLD
MUST ANTICIPATE
THE
POTENTIAL
OF SYNTHETIC
BIOLOGY.**

RESEARCH HIGHLIGHTS

Selections from the
scientific literature

NEUROSCIENCE

White matter matters in autism

A genetic mutation that causes defects in the insulation of neurons may contribute to the cognitive deficits seen in autism and other neuropsychiatric disorders.

Pratik Mukherjee of the University of California in San Francisco and his colleagues used magnetic resonance imaging to study the brains of 23 children with a deletion in a region of the genome associated with such disorders.

The authors found that in these children, the white matter, which helps signals travel between neurons, was structurally defective. The children scored lower on non-verbal IQ tests (such as visual recognition and memory) than those without the genetic deletion.

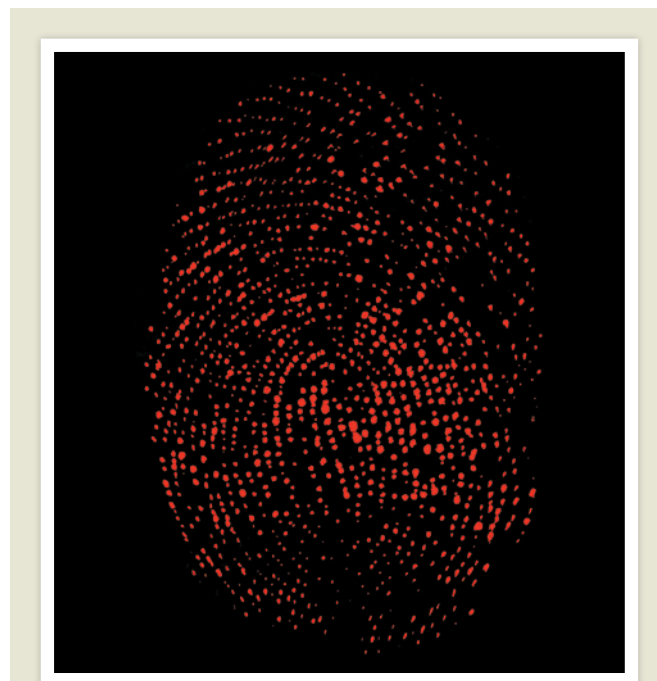
J. Neurosci. 34, 6214–6223 (2014)

CRYOSPHERE

Tipping point for Antarctic melting

Loss of a crucial 'plug' of ice in Antarctica could result in massive sea-level rise, with the discharge of a huge ice sheet.

The Wilkes Basin in East Antarctica contains enough ice to raise global sea levels by 3–4 metres. Anders Levermann and Matthias Mengel at the Potsdam Institute for Climate Impact Research in Germany used topographic data and simulations of ice dynamics to investigate the basin's vulnerability to climate change. They found that if a key plug of ice — equivalent to 80 millimetres of global sea-level rise — were to be lost



MATERIALS

Sweat-pore fingerprint

A polymer that is sensitive to sub-nanolitre volumes of water could be used to map sweat pores for fingerprint analysis.

Jong-Man Kim and Chan Woo Lee at Hanyang University in Seoul and their colleagues developed a thin film made of caesium polydiacetylene that changes from blue to red in the presence of water. The authors found that a fingertip pressed on the blue film produced a pattern of red dots (pictured) that coincided with sweat pores and was visible using an optical microscope.

The sensor technology could be used to identify people by their fingerprint even if only a small part of it was visible, the authors say.

Nature Commun. 5, 3736 (2014)

from the region, the Wilkes Basin ice sheet would be destabilized, creating a self-sustaining discharge of the whole basin over 10,000 years and leading eventually to global sea-level rise on the metre scale.

However, the ice sheet would remain stable if the equivalent of less than 60 millimetres of sea-level rise were lost.

Nature Clim. Change <http://dx.doi.org/10.1038/nclimate2226> (2014)

VIROLOGY

Antibodies for Middle East virus

Two independent teams have identified antibodies that can vanquish a deadly virus first reported in the Middle East.

Since it appeared in 2012, the Middle East respiratory syndrome (MERS) coronavirus has caused disease in at least 261 people and killed nearly

100 in Asia, North Africa and Europe as of 26 April, according to the World Health Organization. There are no vaccines or drugs for the virus.

Wayne Marasco at the Dana-Farber Cancer Institute in Boston, Massachusetts, and Ralph Baric at the University of North Carolina in Chapel Hill and their team identified seven human antibodies that recognize different parts of a key MERS coronavirus protein and prevent it from recognizing and entering human cells in culture.

Another team, led by Linqi Zhang at Tsinghua University in Beijing, identified two more neutralizing antibodies. The molecules recognize different portions of the same MERS coronavirus protein and work together to block the virus's ability to infect cells.

If the antibodies prove to be safe and effective for human use, they could be administered to prevent or treat MERS coronavirus infections, both teams say.

Proc. Natl Acad. Sci. USA <http://doi.org/smr> (2014); *Sci. Transl. Med.* 6, 234ra61 (2014)

METABOLISM

Hunger marks offspring's genome

Male mice that are undernourished while in the womb sire offspring with an increased risk of metabolic disorders such as diabetes.

Josep Jiménez-Chillarón at the Sant Joan de Deu Hospital in Barcelona, Spain, and his colleagues underfed pregnant mice and found that male offspring had a chemical modification on the *Lxra* gene, which regulates fat metabolism in the liver.

This 'epigenetic' change was transmitted through the offspring's sperm to the next

J.-M. KIM ET AL.

CLAUDIUS THIRIET/BIOPHOTO/ARDEA

generation. It was also specific to certain types of tissue, such as the liver and skeletal muscle. Both generations showed reduced expression of genes that synthesize lipids in the liver. This makes them prone to developing glucose intolerance, which can lead to diabetes.

The results suggest that the epigenetic effects of maternal malnutrition affect both the offspring and the next generation.

Cell Metab. <http://doi.org/sms> (2014)

MICROBIOLOGY

Drug resistance from manure

Dairy-cow manure that is spread on farmland could also be spreading a wide variety of genes encoding antimicrobial resistance, reports a Connecticut-based team.

Jo Handelsman at Yale University in New Haven and her team analysed DNA from five samples of manure from dairy cows (**pictured**) and found 80 genes from a variety of bacteria that encode proteins for resistance to a wide range of antibiotics. Among them was a new group of genes that are specific to cow manure and confer resistance to the antibiotic chloramphenicol, which is used clinically. The protein sequences deduced from the 80 genes were, on average, only 50–60% identical to known amino-acid sequences.

Cow manure is commonly used as a crop fertilizer, and this could be one way in which

resistance genes disperse beyond the dairy farm, the team suggests.

mBio 5, e01017-13 (2014)

PHYSICS

Heavy neutrino may be dark matter

‘Sterile’ neutrinos could be the source of a mysterious signal believed to be coming from dark matter in galaxy clusters.

Standard neutrinos are thought to be too light to be candidates for dark matter. But Kevork Abazajian of the University of California, Irvine, calculated how the hypothetical sterile neutrinos, which theories suggest could have larger masses than standard neutrinos, might decay to produce X-ray emissions such as those seen earlier this year by the XMM-Newton and Chandra X-ray space telescopes.

By factoring in the particles’ interactions in the early Universe, Abazajian shows how sterile neutrinos could also be better than other dark-matter models for explaining how mass is distributed in the Milky Way.

Phys. Rev. Lett. 112, 161303 (2014)

MARINE BIOLOGY

Sea life booms in hot climates

Deep-sea ecosystems are sensitive to global temperature variations on shorter timescales than previously thought.

Moriaki Yasahara at the University of Hong Kong and his colleagues took samples of a fossil-rich sediment core from the northeastern Atlantic Ocean to study variability in the abundance of deep-water species over the past 20,000 years. Focusing on a scale of decades to centuries, the researchers found that abrupt glaciations such as the Younger Dryas cooling period some 13,000 years ago led to significant increases

SOCIAL SELECTION

Popular articles on social media

ENCODE debate revived online

In the social-media age, scientific disagreements can quickly become public — and vitriolic. A report from the ENCODE (Encyclopedia of DNA Elements) Project consortium proposes a framework for quantifying the functional parts of the human genome. It follows a controversial 2012 *Nature* paper by the same group that concluded that 80% of the genome is biochemically functional (*Nature* 489, 57–74; 2012). Dan Graur, who studies molecular evolutionary bioinformatics at the University of Houston in Texas and is a vocal ENCODE critic, weighed in on this latest report. ENCODE’s “stupid claims” from 2012 have finally come to back to “bite them in the proverbial junk”, Graur wrote on his blog. The targets noticed. “Some people seek attention through hyperbole and mockery,” says the report’s first author Manolis Kellis, a computer scientist at the Massachusetts Institute of Technology in Cambridge. “We should stay focused on the issues.”

Proc. Natl Acad. Sci. USA 111, 6131–6138 (2014)



Based on data from altmetric.com. Altmetric is supported by Macmillan Science and Education, which owns Nature Publishing Group.

➔ **NATURE.COM**
For more on popular papers:
go.nature.com/9hsjcw

in biodiversity on the ocean floor within around 100 years. Weakening ocean-circulation patterns during these surface-cooling events resulted in warmer deep-ocean waters.

This high sensitivity to rapid temperature variation shows how marine biodiversity could respond suddenly to human-induced climate change this century, the authors say.

Global Ecol. Biogeogr. <http://doi.org/sk7> (2014)

AGRICULTURE

Maize vulnerable to drought

Yields of maize (corn) have increased in the central United States since 1995, but so has the crop’s sensitivity to drought.

David Lobell from Stanford University in California and his colleagues studied yields of US maize (**pictured**) and daily weather data for between 1995 and 2012, focusing on a measure of atmospheric dryness called vapour-pressure deficit. The authors



found that yields have been increasing overall, but in years of high vapour-pressure deficits, maize grew more slowly. The team says that the practice of packing plants close together could be boosting this sensitivity.

Climate change could have larger negative effects on US maize crops than previously thought, because vapour-pressure deficit in the region is predicted to rise as the climate warms, the authors say.

Science 344, 516–519 (2014)

➔ **NATURE.COM**
For the latest research published by Nature visit:
www.nature.com/latestresearch



SARAH M. GOLONKA/TETRA IMAGES/CORBIS

SEVEN DAYS

The news in brief

POLICY

Rising resistance

The world is heading for a “post-antibiotic era” in which drugs against common infections will cease to work, warns the World Health Organization. On 30 April, the agency called for a global surveillance system to track how antibiotics are being used and where resistance is cropping up. See page 141 for more.

UK animal research

In a bid to improve transparency, the UK government is considering reforming a law about animal research conducted by the nation's institutions. Current rules prohibit the Home Office from releasing any information that could be considered confidential, such as on the type of animal research or the institutions involved. On 1 May, the government launched a consultation on its proposal to repeal this blanket ban and replace it with a rule that protects information about “people, places and intellectual property”. See go.nature.com/zijvk2 for more.

UK spaceport

The UK government is looking into building a commercial spaceport by 2018. The move, announced in a report on 30 April, is part of a raft of measures designed to boost the UK space industry. A committee will consider locations for the port and review regulation reforms to enable commercial space flight in the United Kingdom.

Reef heritage

The Great Barrier Reef, which lies off the coast of Queensland in Australia, narrowly missed being listed as ‘in danger’ by the committee that manages the list of world heritage sites.



LILY CLARKE/CONSERVATION INTERNATIONAL

More protection for Coral Sea

One of the world's largest marine conservation areas has been created in the Coral Sea by New Caledonia, a French territory more than 1,000 kilometres east of Australia. The government of New Caledonia said on 23 April that a management plan is in development

for the 1.3-million-square-kilometre marine park, but exactly how much of the area will be off-limits to fishing is not yet clear. Australia has already protected large swathes of the Coral Sea, which lies off its eastern coast and is connected to the Pacific Ocean (see *Nature* **495**, 155; 2013).

The preliminary decision taken by the United Nations Educational, Scientific and Cultural Organization committee on 30 April is expected to be finalized at a meeting in June. The committee acknowledged that Australia has taken some measures to manage the reef, but expressed concern about developments such as a planned dumping of dredged material. It warned that the reef could still make the list in 2015 unless it is further protected.

RIKEN paper chase

Pressure has intensified on a Japanese committee investigating misconduct in two stem-cell papers after whistleblowers accused three panel members of problems with their own work. This follows the panel's judgement

of misconduct last month for Haruko Obokata, of the RIKEN Center for Developmental Biology in Kobe (which she is appealing against), and last week's resignation by the panel's head over alleged problems with two of his papers. Last week, RIKEN also made public that on 25 April, the institute's president Ryoji Noyori told all 280 principal investigators to examine their papers; a RIKEN spokesperson notes that, collectively, these researchers publish more than 2,500 papers each year. See go.nature.com/de7dxu for more.

30 April. Mills is former chief executive of Osiris Therapeutics in Columbia, Maryland — the first company to commercialize a stem-cell drug. His appointment is seen as a bid to help the institute to deliver stem-cell treatments. Mills will begin in the post on 15 May, replacing reproductive biologist Alan Trounson, who announced his departure last October. CIRM is one of the world's largest stem-cell funding agencies.

Society fellows

The Royal Society, Britain's national science academy, announced on 1 May the election of 50 new fellows, including England's chief medical officer Sally Davies. The society also announced ten new foreign members, among them Nobel physics

PEOPLE

Stem-cell head

The California Institute for Regenerative Medicine (CIRM) announced Randal Mills as its next president on

NASA laureate and former US energy secretary Steven Chu. Just seven (14%) of the newly elected fellows are women, leaving the Royal Society trailing behind the US National Academy of Sciences, which two days earlier announced 84 new members, of whom 21% are women. See go.nature.com/dehmq9 for more.

FUNDING

Research red tape

Researchers at US institutions are overloaded with administrative tasks that are hampering their ability to do science, warns the National Science Board in Arlington, Virginia, in a report published on 1 May. The report surveyed thousands of federally funded scientists. It found that the most burdensome requirements include writing research grants and reporting progress. "Excessive and ineffective requirements take scientists away from the bench unnecessarily and divert taxpayer dollars from research to superfluous grant administration," says Arthur Bienenstock, chairman of the task force that produced the report.

FACILITIES

New-look spacesuit

Martian-bound astronauts may one day sport spacesuits similar



to the prototype pictured. The bright blue styling unveiled by NASA on 30 April was the public's favourite, winning 233,431 votes (nearly two-thirds). Unlike previous generations of designs, this suit includes a rigid upper torso to protect the astronaut during ventures out onto the Martian surface; it also features more flexible hip and shoulder joints. The suit is expected to be ready for testing by November. It will be put through its paces in the gigantic pool in Houston, Texas, that NASA uses to train astronauts for spacewalks. Trips to Mars, however, remain decades away.

RESEARCH

Element 117

Scientists have created atoms of the superheavy element 117, they announced on 1 May, confirming the element's discovery four years ago by a different group. The latest

experiments were conducted at the GSI Helmholtz Centre for Heavy Ion Research in Darmstadt, Germany (J. Khuyagbaatar *et al. Phys. Rev. Lett.* **112**, 172501; 2014). The atoms, which were created by smashing together lighter elements, were first glimpsed at the Joint Institute for Nuclear Research in Dubna, Russia, in 2010. Official recognition of the discovery, and a name, await the approval of experts at the International Union of Pure and Applied Chemistry and the International Union of Pure and Applied Physics.

New poxvirus found

A virus related to smallpox has emerged in the Republic of Georgia. Epidemiologist Neil Vora of the US Centers for Disease Control and Prevention in Atlanta, Georgia, and his colleagues isolated a virus from two cow herders who became ill after handling sick animals, and discovered that it was not cowpox but a novel type. The researchers reported their results on 30 April at a meeting in Atlanta. Finding a new poxvirus that affects humans is rare, but this is probably because of a lack of surveillance, the team says.

BUSINESS

Sales suspended

The maker of a device used to remove benign uterine

COMING UP

10–13 MAY

The European Society of Clinical Microbiology and Infectious Diseases meets in Barcelona, Spain. The agenda includes antimicrobial resistance and viral infections in pregnancy. go.nature.com/n6wqz8

19–24 MAY

The general assembly of the World Health Organization takes place in Geneva, Switzerland. The meeting will elect a new president and discuss a global strategy to prevent tuberculosis. go.nature.com/myojx8

growths has suspended its sales, following a warning that the device increases the risk of spreading undetected cancer. Johnson & Johnson of New Brunswick, New Jersey, announced the suspension on 29 April, after the US Food and Drug Administration recommended that physicians eschew the device. Known as a power morcellator, it is used to break up growths known as fibroids so that they can be more easily removed during surgery. See go.nature.com/osoyjw for more.

Pharma mega-deal

Pharmaceutical giant Pfizer, headquartered in New York, has been rebuffed after a second attempt to buy London-based drug firm AstraZeneca, a merger that would create the world's largest pharmaceutical company. On 2 May, AstraZeneca rejected Pfizer's improved bid of £63 billion (US\$107 billion). But on 5 May, Pfizer's chief executive Ian Read said that he still intended to pursue a transaction. See go.nature.com/yqgki3 for more.

► **NATURE.COM**

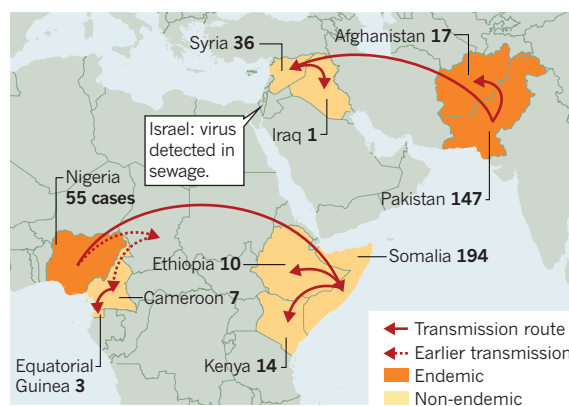
For daily news updates see: www.nature.com/news

TREND WATCH

The spread of wild poliovirus is a public-health emergency that requires a coordinated global response, the World Health Organization declared on 5 May. The agency said that people in countries currently spreading cases abroad (Pakistan, Cameroon and Syria) must be vaccinated before travelling; other nations in which polio is endemic (see chart) should encourage immunization. The measures were invoked under the binding International Health Regulations, last used in the 2009 H1N1 influenza pandemic.

POLIO-SPREAD FEARS

Wild poliovirus is endemic in only three countries, but clinical cases have been confirmed in seven others in 2013–14.



NEWS IN FOCUS

NATURAL HAZARDS Study reveals risks of colonization in avalanche-prone areas **p.142**

POLICY Elections highlight concerns over future of Indian science **p.145**

BIOMEDICAL SCIENCE Universities seek new ways to fund research **p.146**



SYNTHETIC BIOLOGY A tale of two cultures assembling the future of a new field **p.151**

BABU/REUTERS/COREIS



Antibiotics can be bought without a prescription in many countries, and overuse of the drugs is fuelling the evolution of resistant microbes.

PUBLIC HEALTH

Antibiotic resistance sweeping developing world

Bacteria are increasingly dodging extermination as drug availability outpaces regulation.

BY SARA REARDON

By most standards, the increasing availability of life-saving antibiotics in the developing world is a good thing. But, around the globe, overuse of these drugs has created resistant strains of deadly bacteria — and they could be a greater threat in poorer nations than in richer ones, owing in part to a lack of regulation.

Such concerns come amid a stark warning from the World Health Organization (WHO) on 30 April (see go.nature.com/c2l7ry). The world, it says, is poised to enter a 'post-antibiotic' era. And in some places, that era has

already arrived. In Nigeria, for instance, some studies suggest that as many as 88% of *Staphylococcus aureus* infections cannot be treated with methicillin — once a potent weapon against the microbe. The problem also seems to be particularly acute in the emerging economies known as the 'BRIC' states: Brazil, Russia, India and China, says Keith Klugman, an epidemiologist for the Bill & Melinda Gates Foundation in Seattle, Washington.

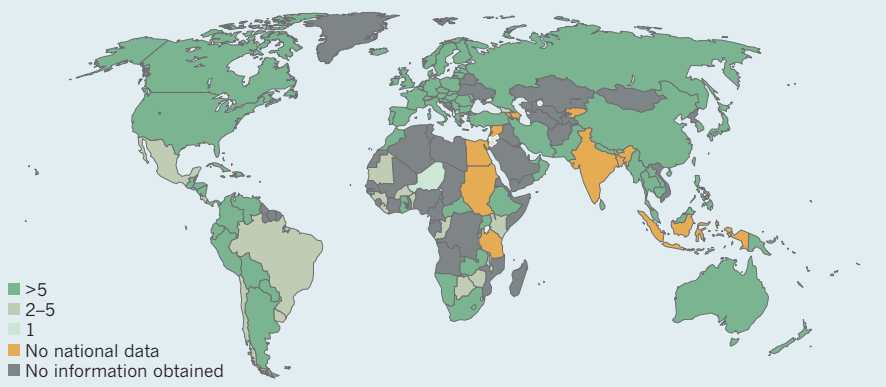
Up to 95% of adults in India and Pakistan carry bacteria that are resistant to β -lactam antibiotics — which include carbapenems, considered to be antibiotics of 'last resort' — according to research by Timothy Walsh, a

medical microbiologist at Cardiff University, UK, that is due to be published in *The Lancet*. By comparison, only 10% of adults in the Queens area of New York carry such bacteria. The spread of resistance is "more than we could have imagined", Walsh says.

And these are only the microbial threats that researchers know about. The WHO's report — which includes the first ever global map of antibiotic resistance — notes that just 129 of 194 member countries provided any national data on drug resistance (see 'Spreading scourge'). And only 22 of those had tracked all of the nine bacteria-antibiotic pairs that the agency names as the greatest threats to public health. That ►

SPREADING SCOURGE

Many countries lack reliable data to track emerging microbial threats, according to the World Health Organization. In large areas of the world, fewer than five antibiotic-resistant bacteria–drug pairs are monitored.



► list includes *S. aureus* and methicillin, *Escherichia coli* and cephalosporins, and *Klebsiella pneumoniae* and carbapenems.

Such problems are compounded by a lack of new antimicrobial drugs in development. This is more than just a problem of future antibiotic supply: pharmaceutical companies' clinical trials of antibiotics were once a valuable source of resistance data for public-health researchers, says Klugman. "Now that we live in a world where there are very few new antibiotics, there is almost none of that support."

As a result, researchers do not know what factors have caused resistance to grow so rapidly in developing nations. For instance, it is not clear to what extent the rise of resistance has been spurred by the use of antibiotics for growth promotion in livestock, or by the release of antibiotics into wastewater by drug-manufacturers in countries such as India.

One likely culprit is a lack of sanitation. In many areas, wastewater from hospitals is poorly filtered, allowing the antibiotic-resistant bacteria that flourish there to escape into waterways. If people drink this contaminated water or practise poor hygiene, the bacteria can spread. "If you get increased antimicrobial use but don't have the infrastructure for infection control, you're setting yourself up for a tsunami of antibiotic resistance," says Klugman.

Other problems include overprescribing or unregulated use of antibiotics. In China, for instance, hospitals and clinics receive financial incentives for prescribing, and antibiotics are overused as a result. Some countries allow pharmacies to sell antibiotics without prescription — and people buy them even for diseases that antibiotics cannot treat, such as malaria. "It's essentially destroying a valuable

resource," says Ramanan Laxminarayan, the director of the Center for Disease Dynamics, Economics & Policy (CDDEP) in Washington DC. Yet insisting that antibiotics be prescribed by physicians in nations where poverty is rife and visits to a doctor are a luxury would mean denying many people drugs that they do need.

Meanwhile, in developed countries, inappropriate prescribing can be a problem if physicians bow to pressure from patients. Training pharmacists and educating the public on correct use of antibiotics could go a long way towards curbing resistance around the world, Laxminarayan says.

Solving these problems will require better monitoring of drug resistance, says Stuart Levy, a physician-scientist at Tufts University in Boston, Massachusetts, who heads the non-profit Alliance for the Prudent Use of Antibiotics. "You can't just go in and scoop up a sample and get results of resistance and think that's the country-wide resistance rate," he says. The WHO report proposes a new global surveillance network, although it is not clear who would pay for it. Levy says that hospitals, non-governmental organizations and local communities could help to track antibiotic usage and test for resistance with inexpensive, readily available screens.

One thing is clear: new types of antibiotic, although sorely needed, cannot offer a solution on their own. They are likely to be expensive and unaffordable in the developing world. More importantly, says Sumanth Gandra, a CDDEP epidemiologist, it is all but inevitable that new drugs will eventually lose their potency, too. "We'll continue the cycle unless we have new drugs and approach the way we use them differently," says Gandra. ■

SOURCE: WHO

NATURAL HAZARDS

Avalanche hotspot revealed

Study of disaster-prone Russian islands underscores perils of colonizing unfamiliar terrains.

BY JANE QIU

Russian writer Anton Chekhov noted in 1895 that the strong blizzards on the island of Sakhalin, once home to a Russian penal colony, were a "friend" to the resident convicts, killing several soldiers. The blizzards are also the main cause of avalanches on Sakhalin and the nearby Kuril islands, and an analysis has now revealed the region to have one of the world's deadliest avalanche records (see 'Snow fall').

The death toll in a 100-year period on the Kurils and Sakhalin — a large island in the north Pacific Ocean — exceeds all

avalanche-related fatalities in the history of Canada or New Zealand. The study highlights the need for proper avalanche assessment and mitigation measures, such as snow fences.

Until now, the extent of avalanches on the islands had remained elusive because of the their divided history, says Evgeny Podolskiy, a Russian avalanche researcher at the French National Research Institute of Science and Technology for Environment and Agriculture near Grenoble, who led the analysis. Fought over by Japan and Russia during the twentieth century, Sakhalin and the Kuril islands — an archipelago that forms part of the 'ring of fire', a band of intense volcanic activity encircling

the Pacific — had no continuous avalanche record. That "has largely hindered the efforts to mitigate avalanche risks", says Podolskiy.

The findings also have implications for the migration of people to mountainous terrains, especially in developing countries, say Podolskiy and his colleagues, who published their study last month in the *Journal of Glaciology* (E. A. Podolskiy et al. *J. Glaciol.* **60**, 409–430; 2014).

Despite a legacy of limited information exchange between Russia and Japan, Podolskiy was determined to piece together a complete history of avalanches on the islands. After talking to colleagues in Japan, where he did

his PhD, Podolskiy found a rare archive. Kaoru Izumi, an avalanche scientist at Niigata University, had a collection of newspaper articles on avalanches that occurred between 1910 and 1945 on the islands, when they were under Japanese control. Izumi had hoped to one day work with Russian scientists.

An astonishing picture emerged from combining and analysing the Japanese and Russian records. Podolskiy, Izumi and their team found that there had been at least 756 fatalities and 238 injuries in 275 avalanches on the islands between 1910 and 2010. The worst avalanche — the deadliest in Soviet history — killed 149 people on Sakhalin in 1945, when it buried a miners' settlement at night.

The study also highlights the danger of small avalanches, says Karl Birkeland, a researcher at the US Forest Service National Avalanche Center in Bozeman, Montana. He says that catastrophic avalanches do not come only in the form of "huge monsters coming down from high mountains over a long distance".

The avalanches that occurred on the islands were mostly on small, gentle slopes with vertical drops of 100–200 metres. It is surprising that this kind of topography could produce such deadly avalanches, says Birkeland.

Mountains on Sakhalin and the Kuril islands are less than 1,600 metres high, and most settlements are concentrated in coastal areas with gently rolling topography. But "the innocent-looking slopes are deceptive", says Podolskiy. The high latitudes and maritime climate mean that the winter is very cold



A train buried by an avalanche on the Russian island of Sakhalin in 2010.

and humid, with heavy snow storms for five months a year.

With a large amount of snow in a wind-swept terrain, "you don't need a big slope to have catastrophic avalanches", says Kouichi Nishimura, an avalanche researcher at Nagoya University in Japan. Early settlements were hidden away from the wind under cliffs with 'snow cornices', large overhangs formed by windswept snow. If these broke away and triggered an avalanche, the destruction could be immense.

The islands' first colonizers "were not aware of the risks and built houses, factories and roads in avalanche-prone areas without proper protection", notes Sergey Sokratov, a snow scientist at the Lomonosov Moscow State University in Russia.

By comparing the avalanche data against population records, the researchers found that periods of high avalanche fatalities — from the 1930s to the 1960s, for example, when the fatality rate was as high as 60 per 100,000 people per decade — were associated with waves of population influx, rapid industrial development and large-scale deforestation. The team showed a similar correlation on the neighbouring Japanese island of Hokkaido, which was also rapidly colonized in the early twentieth century. The trends underscore the importance of

avoiding deforestation, particularly on slopes that are at high risk of snowslides.

Although the decadal avalanche-fatality rate on Sakhalin has since dropped, and was about 1.1 per 100,000 people in 2000–09, "the findings are still very relevant" because a large part of the region's settlements remain unprotected from avalanche risks, says Nishimura, who is in discussions with Russian colleagues to develop mitigation measures on Sakhalin.

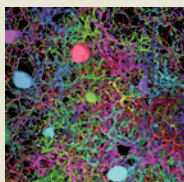
"Historical records are extremely valuable for avalanche mitigation," says Perry Bartelt, a researcher at the Swiss Federal Institute for Snow and Avalanche Research in Davos. In Europe, databases going back more than 300 years are used in planning for disaster mitigation. Such information "is critical for calibrating numerical models to predict how far an avalanche would go at what speed and how much impact it would generate", adds Bartelt.

Sokratov feels that the study has much broader implications as a cautionary tale for people moving to new terrains. "Never rush to build roads and houses without analysing the risks," he says. "With greater awareness and proper mitigation, many lives can be saved from snow avalanches and other natural hazards." ■



MORE ONLINE

VIDEO OF THE WEEK



Wiring of the retina reveals how eyes are able to sense motion go.nature.com/nvnwop

MORE NEWS

- Awareness of speech comes from auditory feedback go.nature.com/ggifuq
- Meditation modulates immune responses go.nature.com/ujgtjs
- Insecticide protects Galapagos birds from parasite go.nature.com/glffiq

NATURE PODCAST



Retina connectomics, whole-Universe simulation, and artificial DNA base pairs nature.com/nature/podcast



Voters queue for the polls in Nagpur, India, as part of the ongoing general election.

POLICY

Indian elections fall flat on science

Researchers concerned about country's competitiveness as most party manifestos neglect innovation.

BY SANJAY KUMAR

As India nears the end of its mammoth general-election process, research advocates warn that whichever party wins, the government must do more than pay lip service to science.

Nearly 815 million people are eligible to vote in the staggered election, which started on 7 April and will end with vote-counting on 16 May. But with two of the three main parties making scant mention of science in their manifestos, researchers fear that Indian science will not get the boost it needs to become internationally competitive.

The manifestos of the ruling Indian National Congress (often referred to as Congress), the Bharatiya Janata Party (BJP) and the new Aam Admi Party (AAP) give mixed clues to what science policies they would pursue.

Congress has supported science in the past, but makes few promises about research now. "Unfortunately, in the value system of both the government and our society at large, science is not at the top" — in fact, it may be at the bottom, says Chintamani Nagesa Ramachandra Rao, chairman of the Indian Scientific Advisory Council to the Prime Minister.

Congress does vow to increase annual science and technology expenditure from the current 0.9% of the gross domestic product (GDP) to

at least 2%. But Raghunath Anant Mashelkar, former director-general of the Indian Council of Scientific and Industrial Research, notes that the 2% spending goal has been mooted by politicians from various parties since 1989. "We have still not crossed 1% for over two decades now, whereas others, such as [South] Korea, China et cetera are rapidly moving towards, or have already moved to, 3% and beyond in the same period." (See 'Broken promises'.)

Of the major parties, the BJP has devoted the most space to science and technology, health, energy and environment in its manifesto. The formulation of the document was chaired by

physicist Murli Manohar Joshi, who was science and technology minister from 1998 to 2004, when the BJP was last in power.

The manifesto talks of reclaiming India's ancient position as a global leader in science, and of creating a productive environment for fundamental research. It promises, for example, to set up world-class centres for nanotechnology, brain research and thorium-reactor technology for India's nuclear-power programme.

Being globally competitive is the main challenge, says Rao. "We have to improve not just the quantity of output, but more importantly the quality of output in science," he says, adding that India currently contributes no more than 1% of the top 1% of global research.

"We have very few islands of excellence in science, sometimes with a very subcritical number of world-acknowledged experts," says Mashelkar.

The BJP manifesto also says that gains made by India during the party's most recent stint in power — when it tested nuclear weapons — have been frittered away by Congress rule. "We will follow a two-pronged independent nuclear programme, unencumbered by foreign pressure and influence for civilian and military purposes, especially as nuclear power is a major contributor to India's energy sector," it promises.

Most opinion polls indicate that a coalition led by the BJP is likely to win the election with prime-ministerial candidate Narendra Modi. Some commentators are uneasy about this prospect. "The rise of Modi will spell a paradigm shift in Indian politics, with a tilt towards militant Hindu ideology or even explicit anti-minoritism, and will impact science adversely," says Pushpa Mitra Bhargava, former director of the Centre for Cellular and Molecular Biology in Hyderabad. In 2004, Bhargava lost a Supreme Court case to halt the BJP government's projects — led by Joshi — to introduce Vedic astrology and Hindu priestly rituals into universities.

The AAP, which is contesting its first general election, has a vigorous anti-corruption agenda. Its candidates include leading anti-nuclear activists and many physicians, but its manifesto is largely silent on science and technology. However, it does focus on management of natural resources such as coal and minerals, which have been at the centre of major corruption scandals in India. The AAP advocates a pivotal role for communities, promising that their consent would be needed to exploit local resources, and that they would share in profits.

Rao says that if India can make careful investments in particular areas of scientific research, it will be able to compete with countries such as China and South Korea in terms of patents acquired and high-quality publications produced.

But he warns: "The competition being very severe, it is my feeling that if India does not catch up in the next 10–15 years, it may become impossible to do so later." ■

BROKEN PROMISES

Successive Indian governments have pledged to boost science spending to 2% of the gross domestic product (GDP), but it has never reached even 1%.



SOURCE: NSTMIS-DST

Universities seek to boost industry partnerships

As drug companies shift focus, academia ramps up its role in bringing discoveries to market.

BY ERIKA CHECK HAYDEN

When Matthew Shair discovered a protein involved in causing blood cancer — and a small molecule that might disrupt that protein — he saw the makings of a new drug. Facing years of costly work to prove its worth, Shair, a chemist at Harvard University in Cambridge, Massachusetts, did not seek funding from the usual sources, pharmaceutical companies and venture-capital firms. Instead, in 2012, he turned to a Harvard programme that supports preclinical research. It provided him with US\$250,000 for preliminary studies in mice and human cells, to determine whether his drug justified undergoing clinical trials.

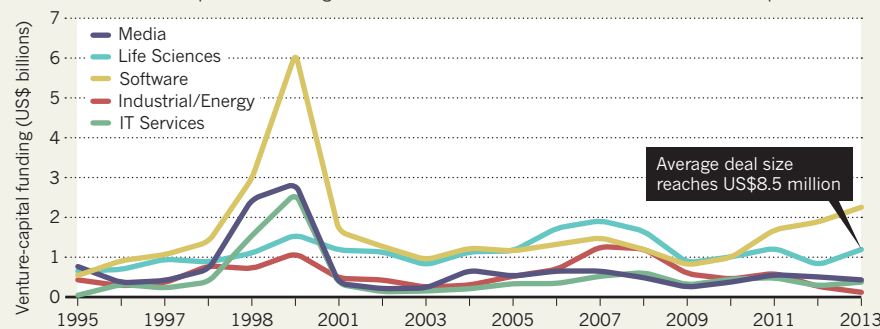
“It gave us the freedom to do the experiments the way we wanted to do them, rather than the way a company might think we should do them,” he says. His experience may soon become more commonplace. Faced with the increasing reluctance of drug and venture-capital firms to fund early development work, academic institutions worldwide are seeking ways to support research that can transform a promising discovery into a marketable therapy (see ‘Capital crunch’).

Funding for early drug-development research has increased over the past year. In April 2013, a \$50-million donation quintupled the size of Harvard’s preclinical programme, which aims to turn a profit and invest that in more grants. And the University of Oxford, UK, last month launched a £1.25-million (US\$2 million) development fund that allows private investors to support the commercialization of discoveries through its business-development unit, Isis Innovation.

Hybrid academic-industrial drug-development schemes are also on the increase. The Karolinska Institute in Stockholm is set to open a \$100-million translational research centre funded by London-based drug firm AstraZeneca in July. A partnership between New York’s Memorial Sloan Kettering Cancer Center, Rockefeller University and Weill Cornell Medical College launched the \$20-million Tri-Institutional Therapeutics Discovery Institute (Tri-I TDI), together with Takeda of Osaka, Japan, in October. And last month, the Scripps Research Institute in La Jolla, California, announced that Johnson & Johnson of

CAPITAL CRUNCH

US venture-capital funding for life sciences has stayed relatively flat over the past decade. In response, universities have developed new funding models to transform discoveries into marketable therapies.



SOURCE: MONEYTREE REPORT (PWC, NVCA)

New Brunswick, New Jersey, would be the first industry partner for its new ‘Scripps Advance’ drug-discovery institute.

Drug companies and research institutes hope that the new programmes will kill two birds with one stone: replenish the pharmaceutical industry’s depleted pipeline of new drugs, and bring money to institutions that face shrinking public funding. “The more quality shots on goal you have, the more quality you get in terms of revenue,” says Scott Forrest, vice-president for business development at Scripps.

Backers of the programmes also hope to avoid the pitfalls that stymied the most recent wave of academic-industrial partnerships in the 2000s. Drug companies outsourced early drug-discovery work to academic researchers,

“We’re seeing a drying up of the basic-research enterprise, which is going to create a problem.”

often by paying for the option to commercialize promising finds. But these arrangements did not yield many new drugs, because undirected basic research

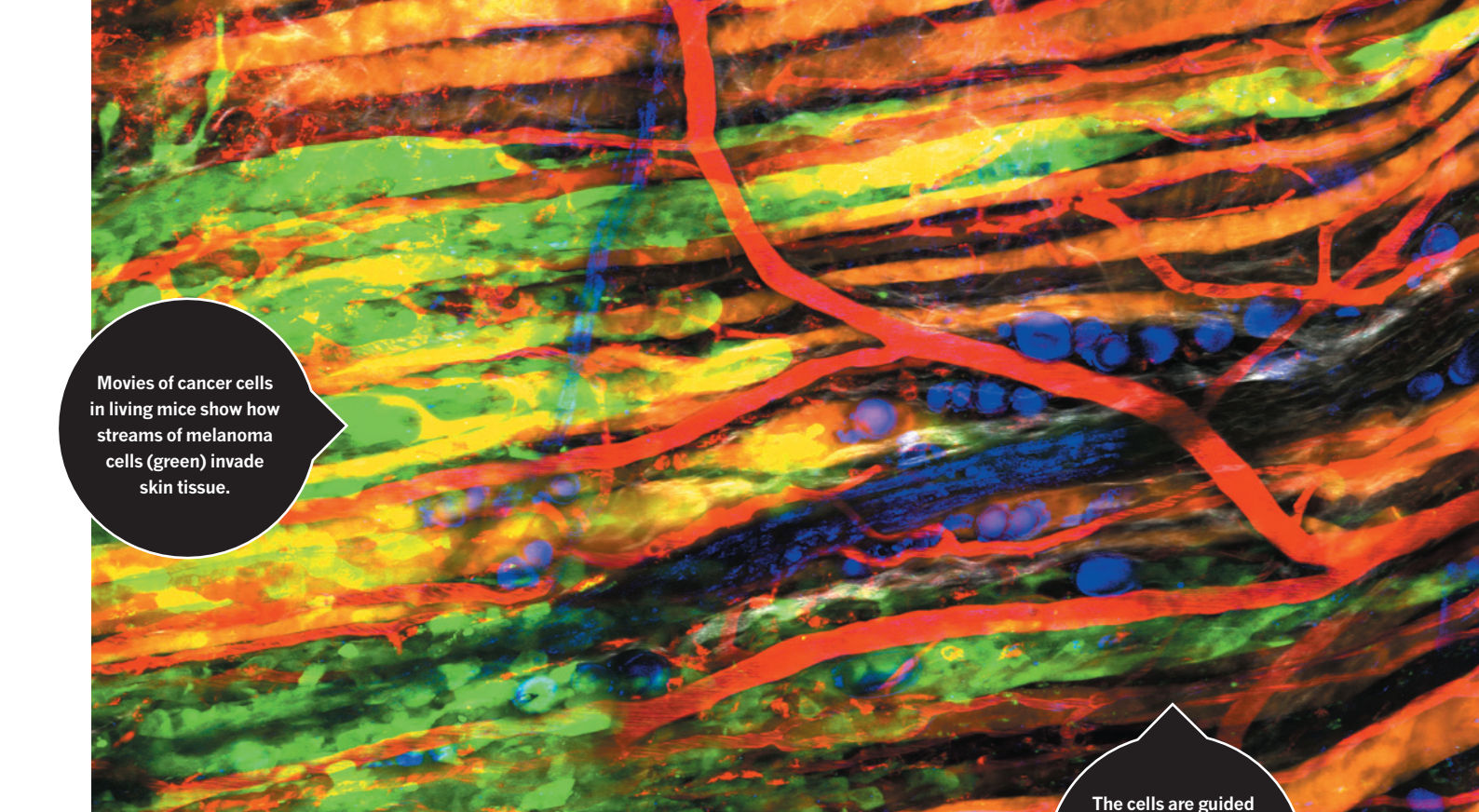
often did not synchronize with industrial goals. The latest partnerships are more focused: they explicitly funnel academic creativity into the pursuit of new therapies. Tri-I TDI, for example, aims to find drugs including those for diseases that industry has eschewed because they are unlikely to generate large profits, such as neglected tropical diseases and rare cancers. Projects will be selected by academic scientists and carried out by Takeda’s scientists at the Tri-I TDI partners’ facilities in Manhattan. “No

one gets to veto a project because it isn’t the right market size,” says David Scheinberg, chair of Memorial Sloan Kettering’s experimental therapeutics centre and a Tri-I TDI board member.

Work at the AstraZeneca–Karolinska centre, meanwhile, is geared towards cardiovascular and metabolic diseases, such as diabetes. The venture is recruiting academic researchers to work alongside AstraZeneca scientists. “We really want to make new discoveries here that, in turn, could eventually possibly be druggable,” says Hans-Gustaf Ljunggren, the Karolinska’s dean of research.

But such arrangements are still vulnerable to market forces. On 2 May, a bid by New York-based drug firm Pfizer to buy AstraZeneca was rebuffed by the company. However, it is not clear how the Karolinska centre would be affected should such a merger eventually go ahead. More broadly, some observers note that if academics and funders continue to shift towards more applied work, there is the risk of a void being created in the basic research that underlies the quest for new therapeutics. Even key funding agencies such as the US National Institutes of Health are shifting towards more translational research. “We’re seeing a drying up of the basic-research enterprise, which is going to create a problem,” says Kenneth Kaitin, director of the Tufts Center for the Study of Drug Development in Boston, Massachusetts.

To many institutions, and to young scientists such as Shair, shaping the development of their discoveries has become even more crucial at a time of scarce resources. “I love having a foot in both sides,” he says. “This is part of the future.” ■



Movies of cancer cells in living mice show how streams of melanoma cells (green) invade skin tissue.

The cells are guided by structures such as muscle fibres (orange), nerve fibres (blue), collagen (grey) and blood vessels (red).

Cancer caught in the act

Methods for monitoring tumour cells in living animals are transforming our view of cancer.

BY CORIE LOK

Mikala Egeblad was blown away when she made her first action film of tumour cells inside live mice. Until then, she had studied samples on microscope slides, where the cells sat still, frozen in time. But seeing them in a living animal brought the cells to life. “You turn on the microscope and look in the live mouse and suddenly these same cells are running around like crazy,” says Egeblad, a cancer researcher at Cold Spring Harbor Laboratory in New York. “It really changed my thinking.”

Increasingly, cancer researchers are embracing the chance to spy on individual tumour cells in their native environment. In studies of static tissue cultures, investigators have to infer what cancer and other cells surrounding the tumour might be doing, and how they might be interacting. Tracking cancer in live animals over time — an approach called intravital imaging — puts those interactions on display, and allows biologists to zoom in on the small number of dangerous cells within a tumour that drive the disease or resist treatment.

The technique is young, and labs are still

working out how best to analyse the gigabytes of video data it generates. But the increasing use of intravital imaging over the past decade has already helped researchers to piece together timelines for key cellular and molecular events, such as the process by which tumour cells sneak into blood vessels. Such clues have yielded new hypotheses about how cancers grow, spread and resist treatment — information that could, for example, eventually enable drug developers to understand why some cancer cells do not succumb to therapy.

And in a video-obsessed culture, the imaging technique holds instant appeal. “When we show our movies, people fall out of their seats when they see how dynamic a tumour lesion can be,” says Peter Friedl at Radboud University Nijmegen in the Netherlands. “It’s a change in perception.”

DEEP DIVE

First used by cancer biologists in the late 1990s, intravital imaging involves focusing powerful microscopes directly onto exposed tissue in a live, anaesthetized mouse. More labs have adopted intravital imaging as technological improvements have made it possible to peer further into tissue — now as many as 20 cells deep — and to tease out fainter signals. A growing library of molecular markers has given researchers the ability to visualize up to eight different kinds of cells and structures,

BETTINA WEGELIN & PETER FRIEDL, RADBOUD UMC

including various immune-system cells and the endothelial cells that line blood vessels. “The markers and the microscopy technology make this a powerful combination,” says Frederic de Sauvage, vice-president of molecular oncology at the biotechnology company Genentech in South San Francisco, California, who has seen the technology in action.

Putting these components together creates a comprehensive picture of cancer as a complex ecosystem of cells that migrate, proliferate and interact. Although cancer researchers have long understood that cells in a tumour are genetically heterogeneous, intravital imaging is revealing how the behaviour of individual cells can also differ. For example, cancer cells may march in single file or collectively as a tight-knit group, depending on the type of tumour and its environment.

One mysterious cellular behaviour that has landed in the sights of these microscopes is that of the macrophage, a type of immune cell that normally engulfs pathogens, removes dead cells and stimulates immune responses. Macrophages can incite immune cells to fight cancer, but more often they boost a tumour’s growth and spread.

Intravital imaging studies showed that macrophages, along with tumour cells and endothelial cells, form a structure that pumps tumour cells into the bloodstream — a key step in metastasis. Working with rodents, researchers led by John Condeelis at Albert Einstein College of Medicine in New York found that when macrophages come into contact with mammary tumour cells, the tumour cells become more invasive, degrading the protein-rich matrix around blood vessels and squeezing between the endothelial cells. Macrophages cause the endothelial cells to lose contact with each other, opening a hole in the vessel wall and allowing tumour cells to stream out of the tissue and into the bloodstream^{1,2}.

Condeelis’s team has shown that this ‘pump’ is present in human breast cancer. The group has also identified three molecular markers, one for each cell type in the structure, that indicate its presence in tumours. In a study³ of 60 people with breast cancer, individuals with a higher density of these pumps in their tumours were more likely to develop metastases in other organs. A start-up company, MetaStat in Montclair, New Jersey, has licensed this prognostic technology and is developing a test that predicts metastatic risk in people with breast cancer. The company hopes to have the test in clinical trials by the end of this year. Condeelis’s group is also working on a probe to identify the pumps using magnetic resonance imaging, avoiding the need to take tissue biopsies from patients.

Others are using intravital imaging to track cancer drugs in the body, and to explore why some drug treatments fail. Cancer biologists typically test the effect of chemotherapies *in vivo* by measuring changes in tumour

growth and size in mice. Intravital imaging gives a more direct view, revealing which cells in a tumour take up the drugs, and whether those cells live or die.

Egeblad and her team have made films of doxorubicin, a naturally fluorescent cancer drug, as it infiltrated mammary tumours in mice. They were surprised by the degree of variability — even within small regions of the tumour — in the amount of the drug that got into the cells, and in the number of cells

“When we show our movies, people fall out of their seats when they see how dynamic a tumour lesion can be.”

that died. One important factor, they found, was the ‘leakiness’ of the blood vessels in the tumour⁴. Mid-stage tumours, which have more porous blood vessels than early- or late-stage tumours, were more sensitive to the drug. Compounds that boost the permeability of vessels could therefore improve the delivery of cancer drugs, suggests Egeblad.

VIEWING WINDOW

To capture films in live mice, researchers were initially restricted to a single imaging session. Ideally, they would like to watch tumours in the same animal over days or weeks to track longer-term changes. Many are adopting a technique that implants a glass coverslip in a frame in the mouse’s skin. These windows, which can provide views into areas including the brain, abdomen and mammary glands, allow investigators to image the same location in the same mouse many times over. The mice wake up after an imaging session and carry on as normal in their cages.

Using the windows, a team led by Jacco van Rheenen at the Hubrecht Institute in Utrecht, the Netherlands, watched colorectal-cancer cells colonize the livers of living mice over the course of two weeks. Newly arrived cancer cells moved within small areas of the organ during the first few days, but by day five they had stopped migrating and were becoming densely packed. The team found that in mice in the early stages of tumour spread, treatment with a molecule that inhibits cell migration decreased the number of metastatic liver tumours that developed later⁵.

As intravital imaging of cancer has matured, the field has moved beyond eye-catching films and has begun to generate quantitative data detailing, for example, the speed and direction of moving cells. Such data allow researchers to construct and refine mathematical models of cell behaviour. These could predict, for example, how tumour cells invade tissues, says Friedl.

But generating such quantitative data is difficult and time-consuming: analysing the movies

can take up to 15 times longer than making them, says Egeblad. Others note that software for quantitative-image analysis is limited, so many labs are writing their own programs.

And intravital imaging continues to pose technical challenges. The technique can access only tissues near the surface, which makes it applicable to just a few tumour types, says de Sauvage. It has also been difficult to integrate intravital imaging with classical tools of molecular biology, such as fluorescent biosensors used

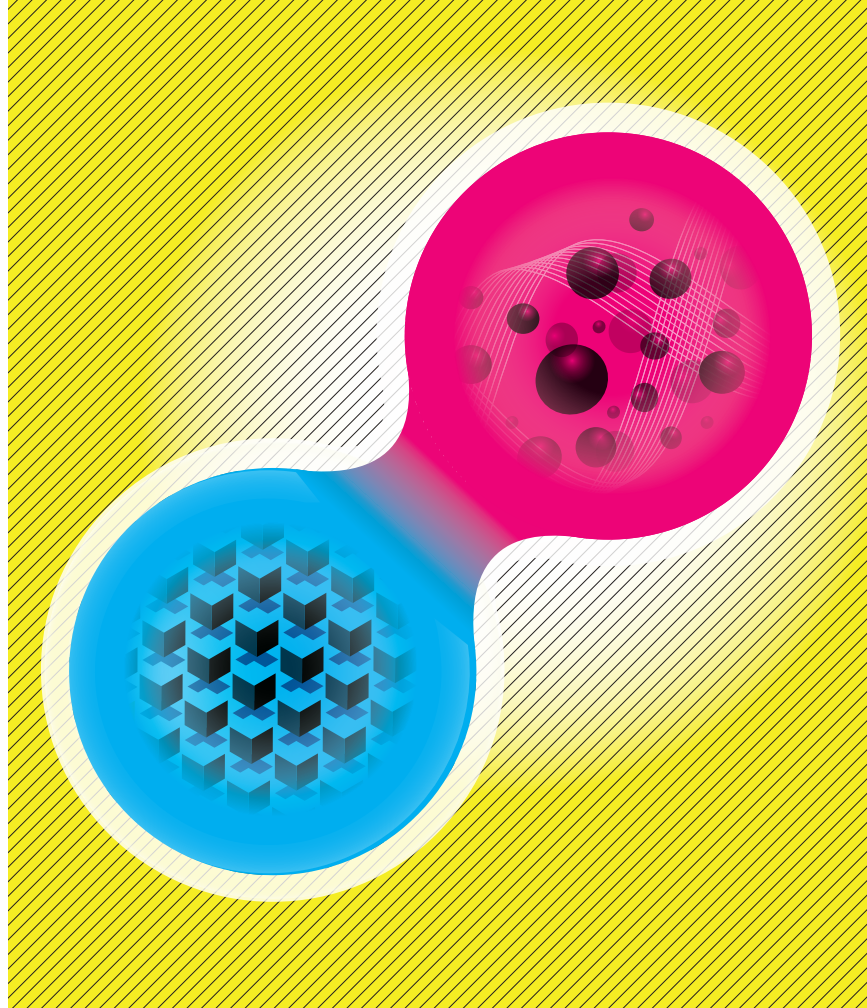
by researchers to see when and where cell-signalling pathways are turned on. Many of those sensors work well *in vitro* — where cells can be manipulated to amplify changes in signalling — but are not sensitive enough to pick up the more subtle changes seen *in vivo*, says van Rheenen. This sort of information might provide clues to the molecular escape routes that allow some tumour cells to dodge the effects of cancer drugs, says Scott Powers, a cancer geneticist at Cold Spring Harbor. “It would be nice to know what’s happening biochemically inside the cells that’s making them do what they’re doing.”

Egeblad is now integrating biochemical and genetic tools in her imaging work. She will soon be launching a new project to trace the history of different subsets of cells in mouse mammary tumours as they grow over several weeks. At the end of the experiment, her team will remove the tumours and sequence the genomes of individual cells. The aim is to link genetic signatures to cellular behaviours, such as rapid growth or drug resistance, in different regions of the tumour. The team also plans to image the activity of key cancer genes in mice as tumours grow.

For Egeblad, the new project is a chance to return to the questions that first drew her to intravital imaging: how do different components of the tumour and its environment co-evolve? Powers says that working with Egeblad and watching her movies helped him to see how the tumour’s environment, not just its genetics, can influence cancer. “How could it not have an impact?” he says. “You’re recording things that haven’t been recorded before.” ■

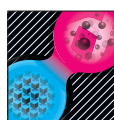
Corie Lok is an editor for *Nature in Cambridge, Massachusetts*.

1. Wyckoff, J. B. *et al. Cancer Res.* **67**, 2649–2656 (2007).
2. Roh-Johnson, M. *et al. Oncogene* <http://dx.doi.org/10.1038/ncr.2013.377> (2013).
3. Robinson, B. D. *et al. Clin. Cancer Res.* **15**, 2433–2441 (2009).
4. Nakasone, E. S. *et al. Cancer Cell* **21**, 488–503 (2012).
5. Ritsma, L. *et al. Sci. Transl. Med.* **4**, 158ra145 (2012).



BEYOND DIVISIONS

Since the birth of synthetic biology almost 15 years ago, the field has splintered into diverse tribes of scientists, all attempting to endow cells with new abilities.



BEYOND DIVISIONS
The future of synthetic biology
nature.com/synbio

Until just a short time ago, the boundaries of biology were set by nature. But those limits are receding. In the past decade, the field of synthetic biology has shattered the idea that living cells must stick to the biomolecular widgets that have evolved naturally. Synthetic biology — a more complex form of genetic engineering — allows researchers to assemble molecular gears on an entirely new scale. From such tools, biological freedom has emerged; even the rule that DNA's alphabet consists of just four letters has been thrown out of the window.

In this special issue, *Nature* surveys the messy landscape of synthetic biology and the various scientific tribes that are pushing it forward. Although the field is young, it is already divided into factions that foresee different futures.

A News Feature on page 152 investigates the schism between those who advocate free-for-all tools, such as open-source repositories, and the groups that seek to protect the fruits of research using legal instruments such as patents (see also Editorial, page 133). Volker ter Meulen follows that theme on page 135, highlighting the threat of excessive regulation in the field. On page 155, experts diagnose the discipline's biggest challenges and suggest solutions through building bridges across scientific and cultural divides. A News & Views Forum on page 166 debates the merits and drawbacks of engineering useful features into cells by fashioning genetic circuits directly or through successive rounds of screening and selection. And on page 168, Daniel G. Gibson and J. Craig Venter describe the research power of a synthetic yeast chromosome built to scramble genes and then probe which combinations can keep cells growing.

Even the most basic building blocks of biology are being remodelled. On *Nature's* website, Denis A. Malyshev and co-authors describe a cell that propagates with unnatural nucleic acids in its DNA (D. A. Malyshev *et al.* *Nature* <http://dx.doi.org/10.1038/nature13314>; 2014). As explored in an accompanying News & Views article, this kind of expanded genetic alphabet might eventually allow the regulatory functions of DNA to be more precisely engineered, or permit new types of amino acid to slip easily into designer proteins (R. Thyer and J. Ellefson *Nature* <http://dx.doi.org/10.1038/nature13335>; 2014). Two of *Nature's* sister journals, *Nature Methods* and *Nature Reviews Microbiology*, are also exploring applications and advances in synthetic biology. Together, the package of reports and commentaries provides a broad overview of how synthetic biology will be building bridges and pushing limits for decades to come. ■



CULTURAL DIVIDE

Synthetic biology is facing a tug of war over whether to patent its discoveries or embrace open-source innovation.

BY BRYN NELSON

A Canadian futurist named Andrew Hessel has an unorthodox idea about how to cure breast cancer. He asks: what if volunteer researchers, working cooperatively from their garages and bedrooms, could rival the efforts of multibillion-dollar pharmaceutical companies?

His crowd-funded venture, the Pink Army Cooperative, is trying to do just that by tapping into open-source tools springing from synthetic biology — an emerging field that designs biological products using engineering principles and a modular approach. Since the cooperative launched in 2009, nearly 600 people have invested in it. The cost to join? A mere US\$20.

This radical idea faces considerable hurdles — but even so, it has attracted plenty of attention from industry groups and the media. The cooperative, launched by Hessel and two co-founders, hopes to start

ILLUSTRATIONS BY THOMAS POROSTOCKY

cell-culture studies this year and is considering a therapeutic trial in dogs.

Currently based at the software-design firm Autodesk in San Francisco, California, Hessel represents an increasingly impatient and outspoken faction of synthetic biology that believes that the patent-heavy intellectual-property model of biotechnology is hopelessly broken. His plan relies instead on freely available software and biological parts that could be combined in innovative ways to create individualized cancer treatments — without the need for massive upfront investments or a thicket of protective patents. He calls himself a “catalyst for open-source synthetic biology”.

This openness is one vision of synthetic biology’s future. Another is more akin to what happens at big pharmaceutical companies such as Pfizer, Merck and Roche, where revenues from blockbuster drugs fund massive research initiatives behind locked doors. For such businesses, the pursuit of new drugs and other medical advances depends heavily on protecting discoveries through patents and restrictive licensing agreements.

Tight controls on intellectual property are necessary to encourage promising medical developments, says the Biotechnology Industry Organization (BIO) in Washington DC, the sector’s dominant trade association. On its website, BIO calls intellectual property “imperative for innovation” around the world. “Societies that protect inventors with patents are the world’s most advanced — scientifically and technologically,” it says.

How synthetic biologists resolve the conflict between open source and patent protection could determine whether the field delivers on its ambitious goal of transforming medicine, agriculture, energy, environmental remediation and other industries through precision engineering. “It’s not just return on investment,” says Linda Kahl, director of the legal programme at the BioBricks Foundation, a non-profit organization in Cambridge, Massachusetts, that advocates for biological engineering in the public interest. “It’s not just commercial applications. It is also about doing good in the world.”

TWO CULTURES

Although its roots extend back to the early twentieth century, synthetic biology started sprouting as an organized field just over a decade ago. In 2003, only 3 peer-reviewed articles listed in Elsevier’s Scopus database used the term synthetic biology; in 2013, more than 800 did. Last year, the field also marked one of its biggest developments. Capitalizing on a discovery by biochemical engineer Jay Keasling of the University of California, Berkeley, the Paris-based pharmaceutical firm Sanofi began large-scale production of a partially synthetic form of the malaria drug artemisinin, which is normally derived from plants (see *Nature* **494**, 160–161; 2013). And more big advances are in the pipeline: at the Pacific Northwest National Laboratory in Richland, Washington, for example, researchers are creating synthetic fungal enzymes that can convert sugars from broken-down plant biomass into fuels and other industrially useful chemicals.

From the start, the field has been an amalgam of disparate influences, each with different cultures of intellectual property. On one side sit software design and engineering, which introduced the idea of encoding desired functions in pieces of DNA and joining together a standardized set of biological widgets, much like bricks or Lego pieces. Software engineers also brought with them the philosophy of sharing their work using open, public registries or only lightly restrictive licensing agreements, such as copyrights.

On the other side sit molecular biology and biotechnology, which supplied know-how about messy and unpredictable biological systems. They also brought the practice of patenting genes, molecules and technical processes. Half of the papers published in *Nature Biotechnology* between 1997 and 1999, for example, were linked to a patent. “They came with different perspectives, different goals, and in some cases, different expectations,” says Andrew Torrance, a law professor at the University of Kansas in Lawrence who focuses on synthetic biology.

Which intellectual-property culture will come to dominate synthetic biology is still unclear. In

June last year, in a case brought by the Association for Molecular Pathology against the company Myriad Genetics, the US Supreme Court ruled unanimously that “products of nature” such as genes and genetic markers are not eligible for patents. But by its very nature, Torrance says, synthetic biology creates DNA that does not occur naturally — and so the court’s ruling explicitly allows such human-designed DNA to be patented.

Legally, therefore, synthetic-biology sequences and techniques can be patented, at least in the United States. But the morals and ethics of doing so are vigorously debated by researchers, companies, lawyers and bioethicists.

Patent advocates say that protecting intellectual property is necessary to spur innovation. In a statement after the ruling against Myriad, Craig Venter, founder and chief executive of Synthetic Genomics in La Jolla, California, applauded the court for making a distinction between naturally occurring and human-derived DNA segments. “These man-made genetic constructs are already being used to create new vaccines, biofuels and nutritional products,” he said. “And the ability to protect this intellectual property is a necessary component of a vital and robust science and biotechnology industry.”

Many synthetic biologists are indeed patenting their work. Writing in *Systems and Synthetic Biology* last year, a group of researchers in Germany documented a trend towards increasing patent applications in the field — particularly in the energy, medical and industrial sectors (D. van Doren *et al. Syst. Synth. Biol.* **7**, 209–220; 2013). Lead author Davy van Doren, an emerging-technologies researcher at the Fraunhofer Institute for Systems and Innovation Research in Karlsruhe, Germany, concedes that the trend is inferred from a limited number of patents and a short time frame — but says that it is consistent with other areas of biology. “We couldn’t find any evidence that patent trends in synthetic biology might be different compared with other domains,” he says.

But open-source advocates argue that patents squelch innovation. This opinion is widespread in start-up companies, non-profit organizations, graduate programmes and the wildly popular annual International Genetically Engineered Machine (iGEM) competition, in which university and school students compete to make synthetic systems that work in living cells. These researchers say that if companies and universities can patent key synthetic-biology tools and building blocks, they can charge hefty fees for others to use them, making it prohibitively expensive to create new products building on those discoveries — especially for start-ups and organizations with few resources.

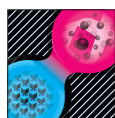
OUT IN THE OPEN

The free-for-all synthetic-biology movement is advancing its aims by assembling public registries analogous to open-source software registries. The iGEM Registry of Standard Biological Parts, the oldest and biggest, contains samples submitted mainly by teams that have entered the competition. By building up a critical mass of components in repositories dedicated to public use, Torrance says, sharing advocates are creating a commons that future innovators can rely on for synthetic-biology building blocks, dubbed biobricks.

Individual parts from these collections can be incorporated into larger and potentially patentable inventions, but theoretically it is harder to patent the basic parts if they are already in the public commons. The iGEM Registry alone is growing by a few thousand parts per year. But there is a big caveat: no one can say with any certainty how many of these parts are themselves entirely free of patent claims. Not all researchers are willing or able to verify that parts that they label ‘open-source’ actually are.

To provide some clarity, the BioBricks Foundation has developed a legal tool known as

**“IT’S NOT JUST
RETURN ON
INVESTMENT.
IT IS ALSO
ABOUT DOING
GOOD IN THE
WORLD.”**



BEYOND DIVISIONS

The future of synthetic biology
nature.com/synbio

the BioBrick Public Agreement: a contract for acquiring standardized biological parts on an open-source basis. Contributors agree not to assert any existing or future intellectual-property rights on a biological part that they have developed, in exchange for a promise by users that they will give proper credit to the developer and abide by security regulations.

Although public registries are rapidly expanding, few have the curatorial capacity to verify that every part works as claimed. “You’re not getting finished, high-quality pieces of DNA in every case,” Torrance says. Some researchers are more blunt. “There’s a lot of crap in there,” says Keasling. To encourage sharing while weeding out the junk, he favours requirements for authors of journal articles to deposit standardized part-source data and descriptions into a repository — provided that the field can agree on a fair set of rules.

Advocates say that the iGEM Registry has improved vastly in recent years, and point to new collections that are emphasizing quality control. The Synthetic Biology Engineering Research Center in Emeryville, California, is developing an Internet interface called the Web of Registries to stitch multiple databases into a linked system that offers uniform information about parts and their legal status. If the field takes off like open-source software, Torrance says, it may attract a do-it-yourself crowd that will help to verify that parts do what is advertised, just as some software enthusiasts currently spend their free time fixing bugs.

Synthetic biologists on both sides of the debate say that few in the field take an absolutist view on patents. Many are instead homing in on the idea of a ‘diverse ecology’ — one that includes both intellectual-property protections and public-sharing agreements (J. Calvert *BioSocieties* 7, 169–187; 2012). Complexity matters here: if the synthetic-biology building blocks are compared to Lego, then in this situation the bricks would be free but a design for a complex rocket ship made of hundreds of Lego pieces would be patentable.

SHARE AND PROTECT

To give an idea of what a robust and commercially friendly open-source regime might mean for synthetic biology, Hessel points to the Linux computer operating system, the open-source platform that became so popular that it is now among the most widely used in the computer industry. Although the base operating system is free, developers have built proprietary businesses onto it, just like biotech companies might be able to incorporate free synthetic-biology building blocks into more sophisticated and patent-worthy systems.

Ginkgo Bioworks in Boston, Massachusetts — which bills itself as the world’s first organism-engineering foundry — is part of an emerging class of synthetic-biology companies that have embraced both public and proprietary models. Among its many projects, Ginkgo is engineering yeast cells to produce chemicals including flavours and fragrances, such as a designer rose extract. The company is not averse to taking out patents on such advanced creations, but it also has a stake in open-source science: Ginkgo sells a \$253 kit for assembling biobricks from iGEM’s Registry of Standard Biology Parts into a multicomponent genetic system. And in 2011, the company agreed to make publicly available one of its engineered constitutive promoters, a DNA regulatory segment that allows a gene to be continually copied into RNA.

“It’s not particularly useful to be patenting individual parts, per se, except in very specific cases,” says Ginkgo co-founder Reshma Shetty. “So I think having the commons available to everyone is a good thing for everyone.”



Although biobricks and other open-access parts have already shown their potential in research projects, advocates say that it is still too early to predict how much they will be accepted into commercial research and development. Ginkgo co-founder Tom Knight, a computer engineer at the Massachusetts Institute of Technology in Cambridge who is often called the father of synthetic biology, says that the field is shifting away from a focus on handcrafted biology, towards a system in which large-scale foundries create standardized parts en masse. If this shift continues, then companies will want to use small parts interchangeably and without complex patent agreements from every manufacturer. Patents will not disappear entirely, says Knight, but they might have a limited role in an industrial context.

Even if the field evolves towards a middle way, the debate will continue. The biotech industry, for one, worries about liability issues associated with free-for-all biological parts. If a publicly available building block is incorporated into a transgenic seed or medical treatment, for example, it is not obvious who is responsible for tracking down its provenance and demonstrating to regulatory authorities that the part is safe.

INNOVATE OR DIE

In the tug of war between patents and open-source registries, a nagging question remains: which mechanism is better at driving innovation? In 2003, the US National Research Council issued a report called *Patents in the Knowledge-Based Economy*, which said that the evidence on how patents affected innovation was still “emergent”. A decade later, the uncertainty persists.

“There are a lot of people looking into this question,” says Torrance. “It is amazing to me that in 2014, it’s impossible to point to a definitive study that indicates that the patent system or the copyright system is a net benefit or a net cost to the economy or to innovation. But that’s where we are.”

It may not matter what the data say, according to Hans Sauer, deputy general counsel for intellectual property at BIO. “For better or worse, we’re just committed to a system that depends on the availability of patents, at least to some extent, for greasing the wheels that put the biotech business model in motion.” In other words, whether patents actually spur innovation may be trumped by the widespread view of entrepreneurs and investors that they are key to minimizing risk in the start-up phase. “These people, rightly or wrongly, all act on their beliefs and their convictions,” says Sauer.

Many in the biotech industry have difficulty imagining a world without patents, he adds. The industry is “a bit spooked” about emerging public hostility to patent rights, and about a legal pendulum that seems to be swinging towards a more restrictive application of patent law, at least in the United States.

The hesitancy is understandable, says Hessel, given the biotech industry’s big investments and long lead times. But a new generation of nimble, leaner open-access types is not bound by such restraints. “What we’re seeing is a kind of transition era, where there’s this new community emerging and it’s in some ways competing intellectually with the current, established industry,” he says. The old guard can go on worrying about downstream investment costs and liability, says Hessel, while he and his Pink Army invent a way to cure cancer.

In March, Hessel spoke to pharmaceutical executives and consultants at the Pharma Summit 2014 in London about the Pink Army Cooperative’s cancer-therapy work. And BIO has invited him to join a panel discussion about emerging trends in biotechnology at the BIO International Convention in San Diego, California, in June. Although the cultural gap between the two camps remains wide, there are signs that the bridge-building has begun. ■

Bryn Nelson is a freelance science writer based in Seattle, Washington.

COMMENT

CITIES Local details are key to success or failure of China's urbanization plan **p.158**

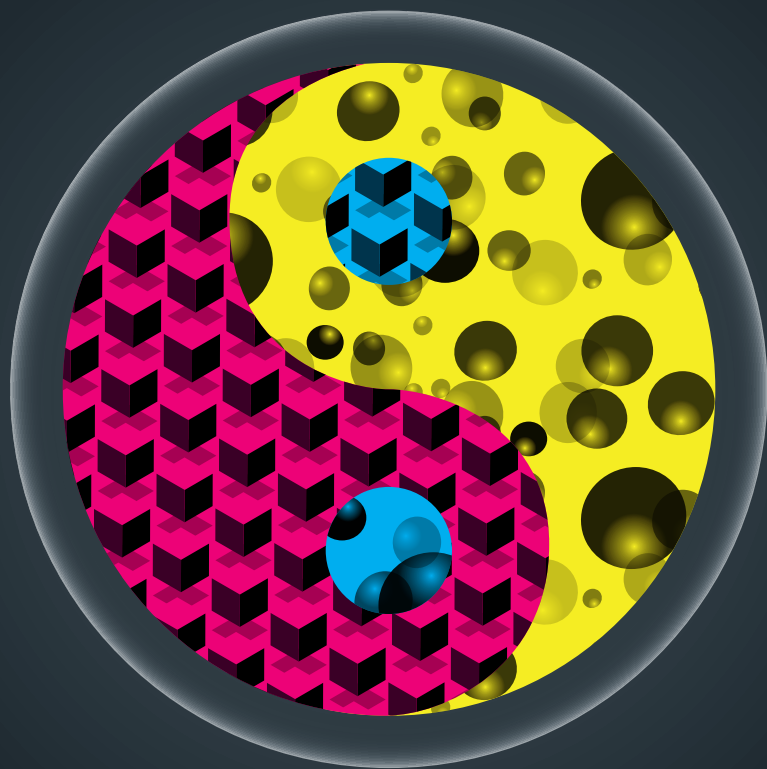
BIOLOGY Embryology has been a contentious discipline since Aristotle **p.161**

MUSIC A composer who works with ecosystem dynamics **p.162**



POLICY Three views on what Indian science needs to thrive **p.164**

ILLUSTRATIONS BY THOMAS POROSTOCKY



How best to build a cell

Experts weigh in on the biggest obstacles in synthetic biology — from names to knowledge gaps — and what it will take to overcome them.

JAMES J. COLLINS

Bring in the biologists

Professor of biomedical engineering, Boston University

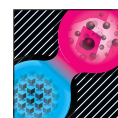
Synthetic biology is often described as bringing together engineers and biologists to build genetic circuits for some useful task. In fact, the field has engaged relatively few biologists. This is holding back its progress. We do not yet know enough biology to make synthetic biology a predictable engineering discipline.

The engineering slant of synthetic biology has brought impressive accomplishments. These include whole-cell biosensors; cells that synthesize antimalaria drugs; and bacterial viruses designed to disperse dangerous, tenacious biofilms.

To design these, engineers are trained to model systems as black boxes, abstractly linking inputs and outputs. They can often control a system with only a limited understanding of it. But synthetic-biology projects are frequently thwarted when engineering runs up against the complexity of biology.

Synthetic biology would benefit greatly from deeper insights into the mechanisms of biological systems. Such approaches have already yielded insights into how organized processes in cells work because of, rather than in spite of, noisy gene expression. Synthetic biology is also informing biology, helping to reveal how a gene product can amplify or inhibit its own expression and so allow cells to flip between stable states. Much more remains to be explored and discovered.

Over the past decade, many engineers have become interested in 'doing biology'. Although biologists need not feel compelled to 'do engineering', biologists and engineers should learn to learn from each other. We must overcome cultural differences and biases. We must move beyond dismissive, tribal comments such as, "Well, she's not ▶



BEYOND DIVISIONS

The future of synthetic biology
nature.com/synbio

► a biologist” or “That’s not engineering”. This requires time, outreach and retraining. But the pay-off will be more contributions that are more meaningful, to both basic biology and biotechnology.

MARY MAXON

Agree on a definition

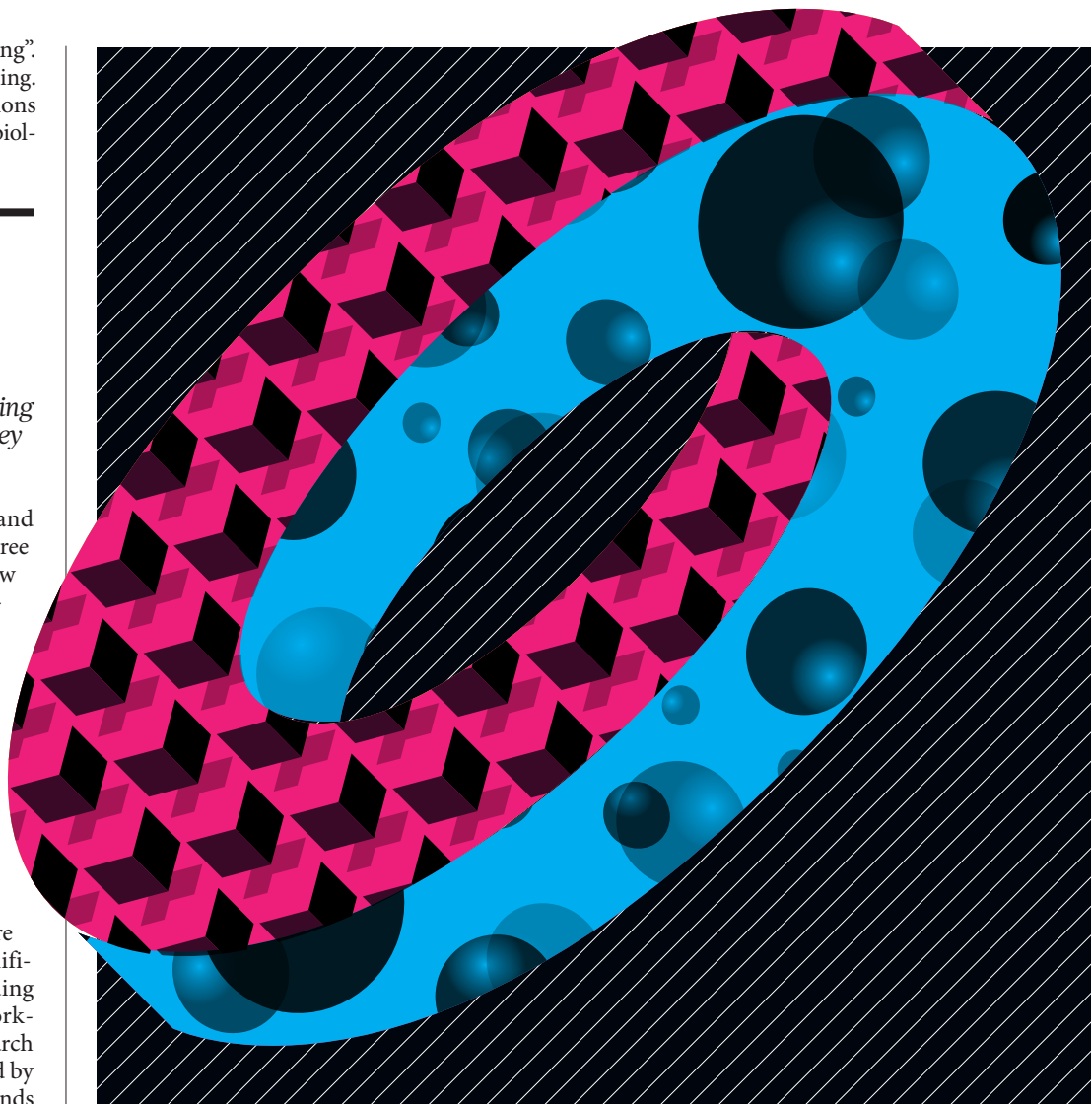
Head of biosciences strategic planning and development, Lawrence Berkeley National Laboratory

Researchers, regulators, consumers and detractors in the United States cannot agree on whether synthetic biology is a new discipline of engineering or an extension of biotechnology. Debates abound over whether geneticist Craig Venter created artificial life using synthetic biology, and whether that term is just another name for genetic engineering. These ambiguities complicate discussions among scientists, hinder policy-makers, impede efforts to fund synthetic-biology research and thwart regulation that might build public confidence.

Emerging technologies that have more precise definitions have captured significant federal support, in terms of funding and policies. For example, the Networking and Information Technology Research and Development programme, created by Congress in 1991, is slated to receive funds of about US\$3.8 billion in 2015, while the National Nanotechnology Initiative, created in 2001, could receive \$1.5 billion.

The federal government already has an interest in synthetic biology. In its *National Bioeconomy Blueprint*, the administration of President Barack Obama included synthetic biology as one of a few technologies for which federal investments could yield economic outcomes through scientific advances. Practical applications spanned health, energy, the environment and food. In November 2013, Congress requested a briefing from the American Chemical Society about how synthetic biology might boost the economy and well-being, having previously asked in the 2012 appropriations bill that the Department of Energy submit a plan to assess the field.

Without a consensus on what synthetic biology is and is not, informed policy is difficult to set. If synthetic biology is the same as biotechnology, it is supported by other programmes, and so does not need a large-scale, federally coordinated effort. Congress is also unlikely to support the field if, as a March 2013 survey found, synthetic biology



is perceived by the public as “unnatural, man-made, and artificial” with as many risks as benefits (see go.nature.com/xlkfmk).

US scientists should follow the example of the UK Synthetic Biology Roadmap Coordination Group, which published its report in July 2012 (see go.nature.com/yneivc). One year later, the UK science minister announced £60 million (US\$101 million) in new support for the field.

ANDY ELLINGTON

Build green ‘bioalchemists’

Professor of biochemistry, University of Texas at Austin

The biggest challenge for synthetic biology is how to extend beyond projects that focus on single products, organisms and processes. Right now, most applications engineer

bacteria that start a synthesis with glucose and turn out biofuels or fine chemicals, such as vanillin or artemisinin. A broader scope could help to build a ‘greener’ economy, in which more organisms make a greater range of chemicals.

The chemical industry is a marvel of efficiency, taking raw materials such as oil and converting them into a wide range of products, including plastics and pharmaceuticals. This is possible in part because feedstocks can be interconverted through various large-scale reactions for which catalysts and processes have been optimized over several decades.

Synthetic biology could unlock the large-scale use of carbon sources from lignocellulose to coal. Synthetic ‘bioalchemy’ would reformat the basic elements of life to take advantage of abundant supplies of formerly rare intermediates such as the nylon precursor adipate, which is used to synthesize antibiotics. Metabolic engineering is already capable of syntheses that use glucose or other standard carbon sources as precursors, but the co-culture of synthetically

modified organisms would make these processes more efficient. The ability to engineer photosynthetic organisms might even allow light to be used as the ultimate energy source and carbon dioxide as the ultimate carbon source.

To bring engineered organisms into the chemical industry, the field must diversify the hosts and acquire a much deeper knowledge of metabolism and ecology so that microbes can be designed to operate on feedstocks much broader than the simple sugars they currently use. Many of the notions of standardization that have so far been the hallmark of synthetic biology will fly out the door. Manufacturing will come to rely on a new class of 'broker' who connects diverse feedstocks and organisms through new chemistries. **Additional input: Hal Alper & Richard Murray**

MARTIN FUSSENEGGER

Make tools for mammalian cells

Professor of biotechnology and bioengineering, Swiss Federal Institute of Technology, Zurich

For synthetic biology to be useful in medicine, more and better tools are needed for work with mammalian cells. The tools that are now standard for bacteria are missing or underdeveloped for mammalian cells. (Similar challenges exist for engineering plant cells.)

To engineer bacteria, researchers can control recombination in genomic segments and also craft large segments of DNA (such as episomes and artificial bacterial chromosomes) to copy and carry out new genetic programs. For mammalian cells, making large chunks of DNA is a challenge, as is getting such DNA to behave properly in the cell. Most mammalian cells hold two copies of their genome, and expression within the genome is controlled in part by epigenetics: chemical markings on DNA and its associated proteins. Mammalian cells also contain specialized structures called organelles, and are programmed to form organized assemblies and to launch cell death.

This means that mammalian cells can provide greater complexity than bacterial cells. Mammalian cells could be engineered to carry out a network of metabolic reactions. Similar to a prosthetic leg or hand, prosthetic networks would perform essential functions. Implanted in the body, these designer cell communities would monitor disease-relevant metabolites in the blood and coordinate the production and release of a therapeutic compound as the body needed it. In mice, such systems have been successful at staving off gout and obesity.

Ideally, prosthetic networks would constantly adjust to the body's requirements, addressing the shortcomings of pills — inflexible units that give patients with variable needs too high or too low a dose. To make designer cells that produce the right amount of drug at the right time, researchers must be able to build better chunks of DNA, decorate them with appropriate epigenetic marks, and get both DNA and epigenetic modifications to persist even when cells divide. No matter how sophisticated the tools, bacteria will not be up to this task.

RON WEISS

Automate efficient design

Professor of biological engineering, Massachusetts Institute of Technology

As synthetic biologists build ever more genetic parts, efficient design has become a major stumbling block. In a system with three components, many of the possible combinations can be explored experimentally, so a design does not need to be perfect. However, once the number of components exceeds five or so, it often becomes impractical to explore the relevant design space completely. Without a much greater effort, synthetic biology will be left with many clever parts, but only limited and inefficient ways to combine them.

When computer programmers create a software application, they do not write every line of code anew, nor do they select each machine instruction by hand. Instead, they reuse existing subroutines and libraries, and leave the job of selecting specific machine instructions to a compiler program.

Synthetic biology needs analogous tools. To make them will require better ways to characterize and tune individual parts across various conditions, such as a cell's growth state. Characterization must go beyond simple 'on' or 'off' categories to reflect a range of activities (see go.nature.com/hr8iu4) and to allow design tools to incorporate modules and parts that can function together. Such parts include insulating devices that prevent modules interfering with each other in undesirable ways or taking too heavy a toll on a cell's resources.

A design tool should hide messy innards to let scientists focus on the logic behind a working system, rather than on the physical construction of its parts. For example, imagine a 'biocompiler' that receives

programming instructions for integrated modules and automatically selects compatible genetic components.

Researchers working on such tools meet annually at the International Workshop on Bio-Design Automation. There are constant improvements in the approach to biological design challenges; still, much more progress is required.

HERBERT SAURO

Capitalize on evolution

Associate professor of bioengineering, University of Washington

Engineering life is not like engineering concrete or silicon. Bioengineers painstakingly craft a design, and a day later it has crumbled in the face of evolutionary selection. Synthetic biologists must learn to address this.

Ideally, designs should function for as long as possible. For that, we can carefully engineer the DNA sequence and use microbial strains that are less susceptible to mutation. We can even construct dual-purpose DNA: read in one direction, it drives the engineered pathway; read in the other, it drives some essential process. Because mutations in the dual-purpose DNA will disrupt the essential process, the cell can survive only if the engineered circuit is also expressed.

Another strategy borrows from nature. Long before human engineers, the bacterial virus phage λ evolved to stay undetected in its host for as long as possible. The virus subtly monitors its host by expressing a single protein at low levels so that the bacterium is barely affected. If the phage detects ill health in the host, it activates a program that ultimately kills the bacterium (and releases new phage particles). Engineers can use similar strategies to kill off cells if they are not performing as engineered.

Conversely, should an engineered organism escape the laboratory, we would like evolution to eliminate it. One could engineer a kill switch or even accelerate evolution so that engineered circuits are eliminated in the absence of some lab-specific conditions. But how these controls would fare in a natural environment is not yet fully understood.

Understanding engineered systems that evolve requires a marriage of disciplines that are not normally associated with each other, and grant panels and agencies are finding it difficult to judge these marriages. On the one hand, engineers may have little time for evolutionary theory; on the other, evolutionists may have little interest in engineering. But scientists must consider both together to get the most from synthetic biology. ■

"A design tool should hide messy innards to let scientists focus on the logic behind a working system."



Development in the city of Jiangyin, which lies close to the Yangtze River near China's east coast.

Realizing China's urban dream

Local implementation and public scrutiny will make or break the government's urbanization strategy, say **Xuemei Bai, Peijun Shi and Yansui Liu.**

China is urbanizing at an unprecedented rate. It is perhaps the greatest human-resettlement experiment in history. Between 1978 and 2012, the fraction of the nation's population dwelling in cities increased from 17.9% to 52.6% (see 'Urban expansion'). If the current trend holds, China's urban population could top 1 billion people in the next two decades.

These are uncharted waters, but China has a plan. In March, the government released the National New-type Urbanization Plan, which sets targets for China's urban population fraction to rise by 1% a year to reach 60% by 2020.

The plan is comprehensive and ambitious. It covers almost every conceivable aspect of urbanization, from rural-urban migration and integration to the spatial distribution of and linkages between cities; sustainable development; institutional arrangements; and implementation. It sets numerical goals (see 'Government targets') and as a guiding principle emphasizes a sustainable and

people-centred approach, paying more attention to welfare and well-being — a significant and positive shift from the current economic focus on land development. It also aims to rectify existing problems associated with the rapid urbanization in the past three decades.

The right national strategy is necessary. But it is not sufficient. It is local practices that will make or break China's urbanization plan.

GOVERNMENT GOALS

President Xi Jinping and Premier Li Keqiang see sustainable urbanization as a policy priority for three reasons.

First, urbanization and industrialization are the two engines of modernization and economic growth. High-income countries typically have high urban population proportions — 70–90% for countries in the Organisation for Economic Co-operation and Development. By comparison, China's urbanization lags behind its industrialization. Research shows that urbanization does

promote economic growth in China¹, even though the causal link between the two has been questioned².

Second, to keep its economy growing, China needs to shift from its current reliance on exports by spurring domestic demand. Urbanization enhances domestic consumption: per capita income in China is more than three times higher in urban than in rural communities.

Third, urbanization can catalyse regional development. With less than 0.1 hectares of arable land per farmer, rural China has little room to boost income growth by expanding agriculture. Urbanization will reduce the poor agricultural population and the number of smallholder farms, and open up possibilities for accelerated rural development through large-scale, modernized farming practices.

Development is at the core of these rationales, and the government aims to do it in a sustainable way. Building and operating growing cities requires materials and energy, but the demand for some of these will exceed the country's supply capacity³. And daunting environmental and social challenges caused by past urbanization need to be avoided.

China aims to guide the development of urban areas according to their resource and environmental carrying capacity; improve land, water and other resource use efficiency; stabilize the employment of migrating workers; and promote green and low-carbon urban construction and operation. Its 12th Five-Year Plan for 2011–15 also aims to reduce construction land use per unit of gross domestic product (GDP) by 30%, energy intensity by 16% and carbon dioxide emission intensity by 17%.

There is, of course, a catch: if the income of new urban residents matches that of existing ones, the consumption that will fuel China's economic growth will present sustainability challenges to the country and other parts of the world. For example, the rise in CO₂ emissions that would result if China's household consumption patterns converge with current US levels could not be offset by efficiency gains alone⁴. Promoting low-carbon lifestyles is mentioned in the plan only in passing and without concrete measures or targets for how they might be achieved.

POLICY CHALLENGES

To live the urban dream, China faces three policy challenges: land, people and the environment.

In the past decade, the urban built-up land area in China has grown by 78.5% — faster than its urban population, which grew by 46%. The rapid pace of landscape urbanization is driven by Chinese cities' over-reliance on income from land release¹, which can amount to 70% of their total revenue.

Land-release income increased by a factor of six between 2005 and 2011.

About half of urban growth has been at the expense of arable land, raising concerns about food security. To curb the loss of agricultural land, the central government has introduced strict regulations, mandating that cities offset developed land by securing the same amount of arable land elsewhere or reducing the area of land used for rural housing.

These policies have had limited success, and in some cases, adverse consequences. In a desperate search for non-agricultural land to develop, cities have reclaimed wetlands and lakes and converted pristine mountains. At the same time, developed land is not being used efficiently — attention quickly shifts to the next development project.

The flow of people from rural communities to cities raises problems. Migrating peasant workers, currently estimated at 250 million, are prevented from integrating into urban communities by the household registration system. This restricts the official changing of citizens' permanent residence, and effectively limits those with rural registration from gaining access to urban medical insurance and other services. Migrants typically live in 'urban villages', often former villages or new shanty towns on the edge of a settlement, without adequate social security, health care or schools⁵.

Although forbidden, 'forced urbanization' is widespread. Rural residential lands are being expropriated with insufficient compensation and inhabitants forced into newly developed apartment buildings. Peri-urban communities — those that straddle urban and rural areas — that are affected by expansion often become more vulnerable both socially and economically, with a sudden loss of traditional livelihoods such as farming and fishing, and without adequate skills and access to other jobs.

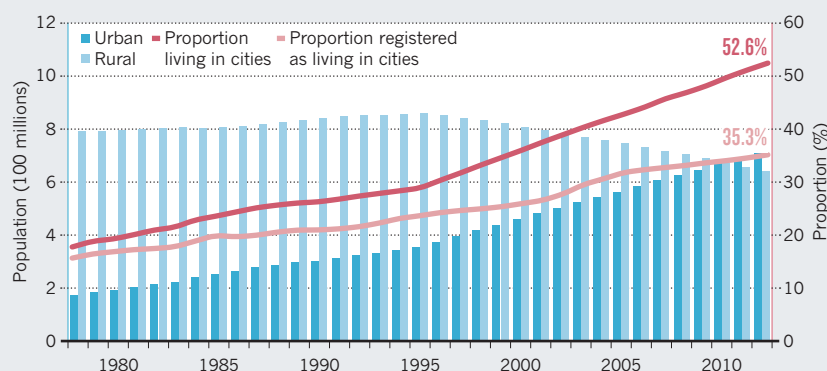
Neglecting migrants' needs has enabled Chinese cities to grow at low cost. Now this strategy is backfiring. For example, the lack of security and sense of belonging of incoming workers in cities makes them reluctant to give up their rural land quota, which is left uncultivated. They invest their savings to build bigger houses back in the villages, often on agricultural land. Urban crimes and protests for appropriate pay by migrant workers are rising.

Although rural populations are smaller than urban ones, and falling, rural residential land use has increased rapidly. Four times more rural land than urban is being used for construction of all sorts, often inefficiently. More than 7.58 million hectares of potential arable land could be realized by optimizing rural residential land use — clearing up discarded residential plots⁶, for example. Integrating migrating workers into

URBAN EXPANSION

The proportion of China's population living in cities has risen steadily since the 1970s (1). Workers moving from rural areas to cities are prevented from integrating, however, by a household registration system that restricts them from officially changing their permanent residence. Urban populations concentrate in the eastern part of the mainland (2).

1 Relocation trends



2 Mainland cities



urban life — for instance, by guaranteeing access to education and providing water and sanitation — has costs, but is essential to achieving the state's policy goals and maintaining social stability.

Urbanization is also linked to environmental and health problems⁷. About two-thirds of Chinese cities are short of water, exacerbated by severe surface and ground water pollution. Air and water quality in cities and the surrounding rural areas are deteriorating, with increases in respiratory diseases and cancer rates⁸. In 2013, Beijing hit a high for the number of days with smog pollution — 46 of the first 100 days of the year. Illegal discharge of industrial wastewater, which contains heavy metals and other pollutants, around cities is blamed by some for more than 400 'cancer villages', in which cancer diagnoses and deaths are sometimes 20–30 times higher than the national average⁹.

Most city development has been in the eastern part of the country, particularly

along the coast. The procession of urban and industrial development inland and westwards, to spread manufacturing across the country and reduce regional disparities, is clearly articulated in the government's strategy. But this could bring polluting industries into regions where fragile ecosystems are providing crucial services to the nation. Pollution in western China, for example, which is home to the sources of the Yangtze, Yellow and Lancang rivers, could have profound effects downstream.

Environmental management strategies that stop at cities' administrative boundaries are no longer effective, as epitomized by the frequent regional smogs. Officials need to identify good ways to collaborate across city and regional boundaries.

LOCAL IMPLEMENTATION

The National New-type Urbanization Plan aims to tackle some of these challenges. The household registration constraints will be removed or relaxed in towns and

GOVERNMENT TARGETS

Main indicators and numerical targets in China's National New-type Urbanization Plan.

Indicator	2012	2020
Urbanization level		
Urbanization ratio (resident population) (%)	52.6	60
Public services		
Proportion of peasant worker's children accompanying parents receiving mandatory education (%)		≥99
Basic social-security coverage for urban and township resident population (%)*	66.9	≥90
Basic medical insurance coverage for urban and township resident population (%)	95	98
Infrastructure		
Public transportation share in total motorized travel in cities with more than 1 million people (%)	45†	60
Public water supply coverage in cities and towns (%)	81.7	90
Proportion of wastewater treated in cities (%)	87.3	95
Proportion of municipal waste decontaminated in cities (%)	84.8	95
Broadband Internet connection capacity of urban households (megabits per second)	4	≥50
Resource and environment		
Per capita urban land use (square metres)		≤100‡
Share of renewable energy consumption in cities and towns (%)	8.7	13
Share of 'green' buildings in new constructions in cities and towns (%)	2	50
Share of prefecture and above level cities that meet the national air-quality meeting standards (%)	40.9	60

*Excludes resident students and people aged 0–16 years. †Data from 2011. ‡Standard is 65–115 square metres for urban built-up area, and 85–105 square metres in newly developed cities. Translated from the *National New-type Urbanization Plan (2014–2020)*.

► cities with populations of up to 5 million. Large investments are to be made to eradicate shanty towns and to provide basic infrastructure in the 'urban villages'.

A new land-management system will increase farmers' property rights, preventing undervalued or illegal land expropriation by local governments. It will restrict rural households from occupying more than one residential plot, putting a stop to rural-sprawl encroaching on arable land. Local governments will be rewarded with central funds for integrating rural populations into the urban system.

The devil, however, is in the detail. Some of these central government reforms conflict with local interests, inviting malpractice. Increased duties of care for migrants and restricted access to land for development will increase the burden on cities and reduce their net revenue.

There are many examples of cities behaving opportunistically. The city of Wuhan, for example, once known for its numerous lakes, has lost 70% of them to reclamation and development in the past 20 years. The Huai River Pollution Control Project, a high-profile, target-driven initiative led by the central government during the 1990s to clean up the heavily polluted river basin, failed when it collided with local income needs. The main measures — shutting down polluting industries and imposing stronger regulations on

industrial pollutions — were relaxed after the target time frame.

To put people at the centre of policies, leaders need to move away from numerical indicators such as GDP growth or urbanization level, which are often pursued at any expense. City leaders need to be encouraged and rewarded for adopting more responsible and sustainable practices, such as taking care of migrating workers.

Local officials must be made more accountable to deter malpractice. Better methods for evaluating personnel and channels for public scrutiny of officials' actions, perhaps on the Internet, need to be established. Serious breaches of regulations must be punishable by law, instead of reprimanded with mere warnings and demotions in the party and the government system.

Some cities are already doing well, exceeding the government's expectations. The city of Zhuhai, for example, has been spearheading the introduction of regulations and legal procedures for making and implementing urban-development plans, promoting public participation in the planning process and aligning urban-development goals with its social, economic, resource and environmental

“Local officials must be made more accountable to deter malpractice.”

context. Zhuhai topped the sustainable-development ranking in the 2013 Urban Sustainability Index of Chinese cities.

WAY FORWARD

Governments must not shirk the responsibilities of protecting farmer's rights; providing proper social security and basic urban infrastructure; and ensuring better environmental quality and health outcomes in cities and surrounding regions. They must address risks and hazards such as flooding and smog.

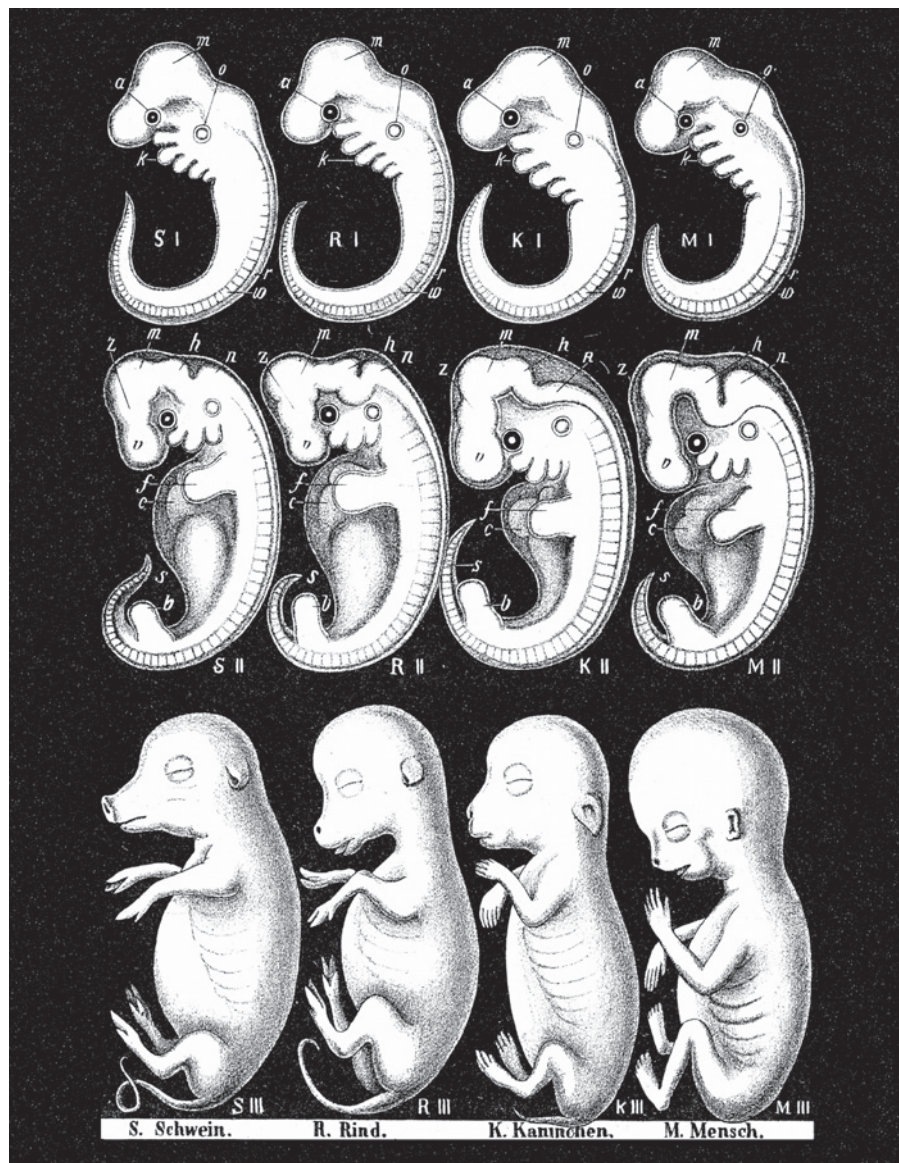
Rapid urbanization is an enormous opportunity to introduce sustainable technologies and practices on a national scale. The latest plan, for example, aims to drastically improve the share of green buildings in new construction from 2% in 2012 to 50% by 2020, by improving green building codes and certificate and accreditation systems, making their adoption compulsory, and developing green construction materials. Efforts must be made to identify innovative practices, build up transferable knowledge and facilitate cross-city learning¹⁰.

High-level policy intervention, if done properly, can be a powerful force for good. But to succeed, a tailored and adaptive policy approach that engages with local social, environmental, economic and cultural contexts and allows local innovation will be crucial.

The ultimate driver of urbanization is — and should continue to be — the aspiration of rural people to lead a better life. And that will be the ultimate measure of success. Urban residents, current and future, must have a say in shaping their cities. They are the ones who will give the verdict on, and live with, the outcome of China's urbanization policy. ■

Xuemei Bai is professor of urban environment and human ecology at the Australian National University, Canberra, Australia. **Peijun Shi** is professor at the State Key Laboratory of Earth Surface Processes and Resources Ecology, Beijing Normal University, Beijing, China. **Yansui Liu** is professor at the College of Resources Science and Technology, Beijing Normal University, Beijing, China.
e-mail: xuemei.bai@anu.edu.au

- Bai, X., Chen, J. & Shi, P. *Environ. Sci. Technol.* **46**, 132–139 (2012).
- Bloom, D. E., Canning, D. & Fink, G. *Science* **319**, 772–775 (2008).
- Shen, L., Cheng, S., Gunson, A. J. & Wan, H. *Cities* **22**, 287–302 (2005).
- Guan, D., Hubacek, K., Weber, C. L., Peters, G. P. & Reiner, D. M. *Glob. Environ. Change* **18**, 626–634 (2008).
- Liu, Y., Lu, S. & Chen, Y. *J. Rural Stud.* **32**, 320–330 (2013).
- Liu, Y., Yang, R. & Li, Y. *J. Geograph. Sci.* **23**, 503–512 (2013).
- Grimm, N. B. et al. *Science* **319**, 756–760 (2008).
- Chen, Z., Wang, J.-N., Ma, G.-X. & Zhang, Y.-S. *Lancet* **382**, 1959–1960 (2013).
- Liu, L. *Environment* **52**, 8–21 (2010).
- Bai, X., Roberts, B. & Chen, J. *Environ. Sci. Policy* **13**, 312–325 (2010).



Ernst Haeckel's 1874 drawings track embryo development in pigs, cows, rabbits and humans.

DEVELOPMENTAL BIOLOGY

Life in flux

Renee Reijo Pera enjoys a treatise tracking the rise of embryology, from Aristotle to cloning and beyond.

Since the advent of *in vitro* fertilization (IVF) in 1978, embryology has rarely left the headlines. From cloning to regenerative medicine to legal wrangles over stem cells and mitochondrial DNA replacement, the science of how and when human life begins has a powerful fascination. In *Embryos Under the Microscope*, Jane Maienschein reminds us that we have always argued about those questions.

Her fascinating tour of them is a balanced

combination of history and science. We track the thoughts of philosophers Aristotle and Descartes; follow the development of modern experimental embryology by scientists such as Frank Lillie; and examine the twentieth- and twenty-first-century focus on understanding the molecular and genetic contribution of the sperm, egg and embryo to the offspring. Through this, Maienschein — director of the Center for Biology and Society at Arizona State University in Tempe, and of

the centre's Embryo Project — interweaves the science of embryology and the many controversies that it continues to spark.

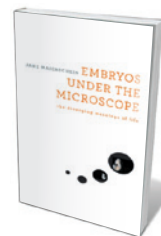
The embryo, the product of the fusion of sperm and egg, was hypothesized from ancient times. Proponents of epigenesis strongly believed that human life begins gradually, and that the individual arises

de novo from potent materials. In the fourth century BC, Aristotle became one of the theory's earliest supporters. He was aware of some 'facts of life', such as that the menstrual cycle was linked to reproduction and that sexual intercourse involved the mingling of male and female fluids. An observer and biologist at heart, Aristotle watched chick embryos develop, seemingly from amorphous materials to a recognizable form that emerged on hatching. As Maienschein makes clear, he instinctively understood that what he observed might be relevant to humans and other animals. Aristotle's ideas illustrate the scientific method at its best, in the face of great unknowns. Without the benefit of microscopy, he put forth concepts regarding the earliest stages of development that survived for more than 1,500 years.

The theory of preformation emerged in the seventeenth and eighteenth centuries along with microscopy. It posited that a pre-formed individual was present at conception but hidden from view, requiring activation. The advent of microscopy allowed scientists such as Caspar Friedrich Wolff and Charles Bonnet to see the earliest stages of development in many animals. This shifted embryology from a largely philosophical field to an experimental one. The earliest stages of species including the chicken, frog, dog and rabbit were examined.

As Maienschein describes, embryology grew by leaps and bounds in the nineteenth century. Karl Ernst von Baer drew elegant pictures of chick-embryo development, and Ernst Haeckel compared the development of multiple organisms, concluding that perhaps "ontogeny recapitulates phylogeny" — development of the individual repeats the evolutionary history of the group or species. In 1895, cell biologist Edmund Beecher Wilson published *An Atlas of the Fertilization and Karyokinesis of the Ovum*, documenting many aspects of fertilization and oogenesis and suggesting that in diverse organisms, the same types of cell may give rise to the same structures.

Concurrently, genetics and developmental biology emerged from embryology ▶



Embryos Under the Microscope: The Diverging Meanings of Life
JANE MAIENSCHIN
Harvard University Press: 2014.

► as scientists sought to explain how development might be linked to chromosomes and heredity, as outlined in studies by Theodor Boveri and Walter Sutton. Further research by embryologist Hans Spemann and others probed how tissues form and what drives cells to develop into different kinds. In recent decades, researchers have been able to genetically modify organisms, track developmental fate and sequence DNA to find naturally occurring variants (insertions, deletions and even single-base-pair mutations) that affect embryo development in flies, fish, frogs, worms, mice and even humans.

Findings in embryology led to a transformation in reproductive medicine with the advent of IVF through the work of Patrick Steptoe and Robert Edwards in 1978. With that came a new focus on the value of human life, from the earliest stages and throughout pregnancy. Concerns burgeoned, leading in the United States to the Dickey–Wicker Amendment of 1996, which prohibited the use of federal funds for research that would create or destroy human embryos. Controversies intensified around cloning technology, beginning in 1997 with the work of Ian Wilmut and his team on Dolly the sheep, and continuing in debates on the ethics of research into human embryonic stem cells and their potential in cell-based therapies for regenerative medicine.

Embryos Under the Microscope does have gaps. It lacks material regarding epigenetic reprogramming — the establishment of DNA and histone-protein modifications in the early embryo that allows embryonic cells with identical DNA sequences to generate diverse cell types — which is a central event in embryology. Yet Maienschein covers broad territory with surprising depth and concision. It seems unlikely that a more readable text will soon emerge to illuminate the journey from theory to observation to ethical considerations in this exciting science.

Lines from T. S. Eliot's 1942 poem *Little Gidding* from *Four Quartets* resonate here: "And the end of all our exploring/Will be to arrive where we started/And know the place for the first time." We have made remarkable progress in our wonderful journey to understand the origins of our development. ■

Renee Reijo Pera is vice-president for research and economic development at Montana State University in Bozeman, and a professor in the departments of cell biology and neurosciences, and of chemistry and biochemistry.
e-mail: renee.reijopera@montana.edu



Daniel Jones in Thetford Forest, UK, the first location for his installation *Living Symphonies*.

Q&A Daniel Jones

Canopy composer

Sound artist Daniel Jones creates self-generating artworks based on human and natural patterns and processes. As he prepares to travel through four UK forests with the installation *Living Symphonies*, a collaboration with artist James Bulley, he talks about music that emerges from ecosystem dynamics, and works inspired by bacterial genetics and social networks.

What is *Living Symphonies*?

It is an ever-changing piece of music that grows in the same way as a forest ecosystem, from the interactions of countless tiny elements. Each reflects the activity of an individual organism, ranging from moss and fungi to deer and birds of prey. The outcome is an organic, emergent symphony comprised of thousands of musical motifs, each portraying a different aspect of the ecosystem.

How do you build each forest installation?

We start with a survey of the animals and plants. We add behavioural insights from ecologists: what times of day is a blackbird active and foraging, how does it move, what are its preferred food sources? We feed this information into a computer model of the ecosystem (programmed in C++ with the Cinder visualization library) that simulates the second-by-second interactions between species. This model is linked to a custom piece of audio software that orchestrates music in real time from a vast

array of motifs representing each species. The music is played through a network of speakers in the canopy and undergrowth.

How does the composition sound?

The motifs for each organism are drawn from fragments that we composed and recorded with orchestral musicians. Our goal is a work with such a great number of interdependent elements, in a nearly infinite combinatorial space, that even we are surprised by the patterns and permutations that arise. This 'emergent' approach runs through all of my work. Emergent phenomena pervade economics, ecology, linguistics and neuroscience; examples include the flocking of birds, and the interactions of neurons that give rise to cognition. By translating the dynamics of a forest ecosystem into music, *Living Symphonies* aims to heighten awareness of the adaptive and often creative behaviours of these complex systems.

Have any animals responded to the piece?

When we ran our first forest prototype last autumn, there was some worry that it would scare off the wildlife. Yet when we

JADE HOFFMAN/NATURE

Correspondence

India: shed the bad science image

Contrary to Mathai Joseph and Andrew Robinson's implication, there are many joys associated with doing science in India (*Nature* **508**, 36–38; 2014).

For example, as a theoretical physicist, I am free to pursue curiosity-driven science, I spend little time writing grant proposals and I do not have to raise money to fund my group members. This fills my foreign colleagues with envy. I suspect that these colleagues, many of whom have made productive scientific visits to India, would rank Indian science well above 166th in the world on a different scale of research quality.

I also disagree that we need “non-resident Indians” to partially staff an “empowered funding agency”: India already has one of the world's largest scientific workforces. To me, this proposal smacks of colonialism, albeit in a new form, perhaps inspired by department closures and tenure abolitions in many Western universities.

Finally, I dislike your portrayal of the Taj Mahal wrapped in red tape to convey the call to ‘Free Indian science’. Thirty years ago a *Nature* cover also featured a Taj Mahal postcard image. That seemed inappropriate to me, even as an undergraduate, for a serious discussion of science in India (see *Nature* **308**, 581–600; 1984). I am now a professor and Indian science has arrived, along with an era of cultural sensitivity — but the iconography remains unchanged.

Shobhana Narasimhan *Jawaharlal Nehru Centre for Advanced Scientific Research, Bangalore, India.*
shobhana@jncasr.ac.in

India: research split harmed universities

One factor that Mathai Joseph and Andrew Robinson do not consider in their discussion on Indian science is the disastrous

effect of uncoupling education and research in the country's universities (see *Nature* **508**, 36–38; 2014).

Several research institutes were created throughout India after it gained independence. These took away good students from the universities and made funding harder to come by, causing a decline in research. In turn, this diminished the number of science faculty members in universities and marred the spirit of education.

This is the opposite result to that predicted by the celebrated Indian spectroscopist Chandrasekhara Venkata Raman. He declared in the 1950s that the institutes would become “mausoleums of science”. He strongly believed that the best research could only be done by the universities (see P. Balaram *Curr. Sci.* **75**, 977; 1998).

Biswa Prasun Chatterji
St Xavier's College, Mumbai, India.
biswaprasun@gmail.com

India: overhaul university teaching

We believe that the quality of education in India is another factor that has been hampering scientific progress in the country since independence (see M. Joseph and A. Robinson *Nature* **508**, 36–38; 2014).

Hundreds of thousands of students graduate from Indian universities each year. However, our own experience in selecting students indicates that many are ignorant of the basics, with underdeveloped reasoning skills and an inability to apply the knowledge they have.

The inadequacy of teaching methods could explain why none of India's universities is among the world's top 200. Learning needs to be more enquiry-based and tutorials more demanding; syllabuses should be overhauled, upgraded and modernized; and, to improve accountability, student-feedback mechanisms

must become more effective.

It is also important to emphasize the value of high-quality teaching skills when recruiting new faculty members, so that the students' technical and reasoning abilities can be fully developed.

Unless India reinvigorates its universities by reforming the teaching–learning process, it will keep failing to generate the workforce necessary to propel Indian science forwards.

Dhruba J. Saikia *Cotton College State University, Assam, India.*
vc@ccsu.ac.in

Rowena Robinson *Indian Institute of Technology, Guwahati, India.*

Aim for balance in Ukraine reports

Your assessment of the scientific implications of tensions arising between Russia and the United States over the Ukraine situation seems to reflect a Western political bias. This could weaken the very research collaborations you seek to encourage (see *Nature* **508**, 149–150 and 162; 2014).

The language you use to describe the Russian regime — for example, “Moscow's violation of international law cannot be tolerated” and “Western observers are stunned by Putin's resolute pursuit of power” — would come as no surprise from, say, US or European media with a clear political leaning. It seems inappropriate from an ‘international journal of science’.

A genuinely international perspective would not overlook the wider political narrative of countries such as China, India and Brazil, who did not join the immediate condemnation of Russia.

Scientific advances in international research environments generally flourish despite cultural and political differences. They may not be so resilient if suspicion and hostility are aroused.

Andrew Isaac Meso *CNRS and Aix-Marseille University, Marseille, France.*
andrew.meso@univ-amu.fr

Help poster sessions to grab attention

Research findings at conferences are now more often conveyed using posters than by oral presentation. Yet the development of posters as a medium has not kept pace with the steady increase in their popularity.

Posters were first used to convey scientific results 45 years ago at the 6th meeting of the Federation of European Biochemical Societies. Sessions were similar to today's, but were shorter and had only about 15 displays per session, so audience engagement was easier.

The worldwide increase in poster presentation as a communication medium should benefit authors and users. Instead, poster authors complain of limited exposure for their work, resulting in a lack of kudos and restricted networking opportunities with other conference participants.

Improving audience circulation, organizing sessions better and providing clear presentation guidelines would help the results on display to capture and hold attention.

Nicholas Rowe *University of Lapland, Rovaniemi, Finland.*
nrowe@ulapland.fi

CORRECTION

The Outlook article ‘Neuroscience: Hardwired for taste’ (*Nature* **486**, S7–S9; 2012) incorrectly stated that L cells manufacture a glucose transporter: it is the absorptive intestinal cells that make this. Also, the article overlooked the contribution of the Shirazi-Beechey lab at the University of Liverpool in confirming sweet and umami receptors in the gut.

FORUM: Synthetic biology

Engineering explored

Synthetic biology involves the creation of biological systems for new applications by modifying and reassembling biological components. Two views are presented here on the best way to engineer these components so that they reliably generate organisms with desired traits.

THE TOPIC IN BRIEF

- The aim of synthetic biology is to predictably bioengineer organisms that perform beneficial functions — from producing antibiotics to purifying contaminated water.
- One approach is rational design, which involves characterizing many biological components to generate a library of modules that can be assembled within an organism to give predictable, reliable outcomes (Fig. 1a).

- An alternative approach is called directed evolution, in which genetic mutations of unknown impact are introduced into a target of interest, generating a library of mutants that is screened for desired characteristics. Iterative rounds of the process produce mutants with optimized traits (Fig. 1b).
- Opinions differ on which of these methods is the more effective.

Rationalizing nature

PAMELA A. SILVER & JEFFREY C. WAY

Rational design is a mainstay of synthetic-biology ideology. What does this term actually mean? As proponents of the rational-design approach, we believe that human designers can use nature to inform the creation of genetic circuits, proteins and even life forms. A human designer can learn from failures, which should ideally inform future designs. By contrast, evolution can be seen as a series of makeshift solutions, in which each change must work at some level for survival — non-working intermediates are not permitted.

Rational design relies on the idea that biological systems are fundamentally modular. The gene was originally defined as the basic biological unit. But the advent of technology allowing manipulation of DNA has revealed tantalizing levels of modularity that extend to many other cellular regulatory elements, including promoters, which control gene expression, and binding sites for the ribosome (the cellular machinery responsible for translation), which stimulate protein production. The dream of the rational designer is to understand these modular parts in sufficient detail to be able to assemble them logically, much as an engineer would build a machine for a certain purpose. There are already success stories — for example, assembly of simple genetic circuits that rely on the existence of two stable

states in a system¹. Recently, we used these principles to design a bacterium that can sense whether or not an animal has been exposed to antibiotics as it travels through the gut².

We are aware, however, that rational design alone is unlikely to be a valid strategy for synthetic biology. Instead, we see it as an overarching approach that will use many tools, including knowledge of physical principles and natural biological systems, computer simulations and more. The synthetic biologist's toolbox will probably contain some level of directed evolution or random variation, but such strategies are often limited by throughput: sometimes animal testing or clinical trials, which can be inherently low throughput, are the most interesting tests of the outcomes.

Ideally, in designing a biological system, synthetic biologists would understand (and apply) all of the phenomena that nature manipulates. This would require a detailed mastery of gene expression, protein structure, enzyme kinetics and so on. But although many of these are well understood, the data are generated by specialists who often do not interact with each other and who frequently do not speak a language that can be understood by bioengineers.

Moreover, most of the scientific literature has been written with the goal of making natural phenomena less mysterious, rather than providing guidance for a bioengineer.

Practical facts and principles are diffusely scattered, and buried in an excess of documentation. One immediate challenge is to consolidate these data into information that engineers can use.

The past 50 years has seen an explosion of biological knowledge, funded mainly at public expense. People might assume that such knowledge is wasted if it cannot be used to rationally design biology for real-world applications. Now is the time to make good on the public's investment.

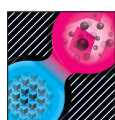
Pamela A. Silver and Jeffrey C. Way are at Harvard Medical School and the Wyss Institute for Biologically Inspired Engineering, Boston, Massachusetts 02115, USA.
e-mails: pamela_silver@hms.harvard.edu; jeff.way@wyss.harvard.edu

Evolving with purpose

FRANCES H. ARNOLD
& JOSEPH T. MEYEROWITZ

Genomes are great books of instructions on 'how to': how to extract materials and energy and convert them into self-repairing, self-reproducing machines; how to function in an extraordinary range of environments; even how to adapt over the long term. Evolution has accumulated a vast library of instructions. We argue that it is also a great tool for writing new ones. The results of directed evolution are rational and predictable, even if the underlying genetic changes follow an uncharted course.

How good are synthetic biologists at writing useful biological instructions? We are getting better at synthesizing long stretches of DNA that can be inserted into a variety of organisms in which they will be read and — sometimes — acted on. But most of our writing merely rearranges passages lifted from the few genomes that we have read. We are plagiarists, poor crafters of original literature. We cannot yet create an enzyme or a biosynthetic



BEYOND DIVISIONS

The future of synthetic biology
nature.com/synbio

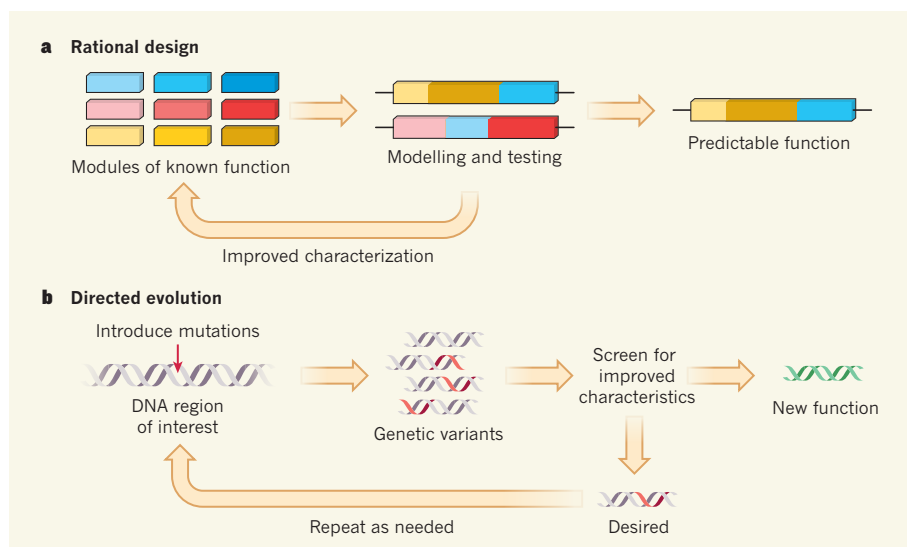


Figure 1 | Two methods for the synthetic engineering of organisms. **a**, In rational design, biological 'parts' (such as genes, gene regulators and proteins) are studied through modelling and testing such that their behaviour is understood. This produces libraries of well-characterized parts — or modules consisting of multiple parts — that can be rationally assembled in cells in different combinations, resulting in predictable functions. **b**, An alternative approach is to harness the power of evolution to direct the design of synthetic organisms. A region of interest (in this example, a gene) is randomly mutated, producing cells that harbour a variety of mutations. The cells are then screened to find those with mutations that result in or improve a desired function. Cells that do not harbour such mutations are eliminated from the pool; those that do can be cultured and then subjected to new rounds of mutation and screening, to achieve a desired end product.

pathway that compares favourably with nature's engineering outputs. Nature's writing is intricate (some say convoluted and opaque), but it is effective. We are just learning to hold the pencil.

The reason is simple: in biology, details matter a lot, and we don't understand the details. Rational design is hard when one cannot even predict the effects of a single mutation in a single enzyme, or the full impact of adding a single gene to the thousands already present in an organism. It is fine to hope that 'modular' biology is possible, but our ham-fisted attempts at assembling biological components usually show that biology is anything but modular. The engineering of systems that tackle real-world problems, such as producing an antimalarial drug or an alternative fuel through a new microbial pathway, involves years of trial and error³. Realizing the potential of synthetic biology requires dealing with the details.

Rational design will not move beyond these problems until our understanding of the details of biology has improved dramatically. Luckily, we do not have to wait. Evolution is a time-tested tool for engineering the details, and we can use it in the lab to circumvent our profound ignorance of how sequence encodes function. Accumulating beneficial mutations over multiple generations is a general algorithm for turning a poor copy-and-paste job into effective writing, and it works at all levels of complexity. From creating a DNA-editing enzyme that excises HIV from the host genome⁴, to optimizing multi-enzyme biosynthetic pathways⁵, directed evolution

has produced results with purpose⁶.

Evolutionary engineering is not incompatible with 'rational' design; in fact, the two are highly complementary. Directed evolution requires a starting design, and the better the design, the easier the evolution. Evolution also requires a rational search strategy. Where

should mutations be targeted, and how many of them? How do we measure success along the way? Directed evolution and rational design are even claiming common ground, for example in mutant libraries that have been designed with input from computational processes, accelerating the evolutionary process⁷.

With a good starting point in hand, evolution is the most direct approach to engineering the biological world, and it is uniquely effective with biological substrates, the products of that same process. Biology is highly evolvable, and we should exploit that feature to the fullest, as we have for thousands of years with everything from rats to racehorses. The writer's best friend is a good editor — the synthetic biologist's should be directed evolution. ■

Frances H. Arnold and Joseph T. Meyerowitz are in the Division of Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, California 91125, USA.
e-mails: frances@cheme.caltech.edu; jmeyerow@caltech.edu

- Gardner, T. S., Cantor, C. R. & Collins, J. J. *Nature* **403**, 339–342 (2000).
- Kotula, J. W. *et al.* *Proc. Natl Acad. Sci. USA* <http://dx.doi.org/10.1073/pnas.1321321111> (2014).
- Keasling, J. D. *Proc. Am. Phil. Soc.* **156**, 283–294 (2012).
- Sarkar, I., Hauber, I., Hauber, J. & Buchholz, F. *Science* **316**, 1912–1915 (2007).
- Bastian, S. *et al.* *Metab. Eng.* **13**, 345–352 (2011).
- Romero, P. A. & Arnold, F. H. *Nature Rev. Mol. Cell Biol.* **10**, 866–876 (2009).
- Trudeau, D. L., Smith, M. A. & Arnold, F. H. *Curr. Opin. Chem. Biol.* **17**, 902–909 (2013).

CLIMATE SCIENCE

The origin of regional Arctic warming

Observational data and modelling show that the rapid warming of the northeastern Canada and Greenland sector of the Arctic over the past three decades has been strongly driven by cooling in the tropical Pacific Ocean. SEE LETTER P.209

JÜRGEN BADER

Over the past 30 years, Earth has become a warmer place. One of the most striking examples of surface-temperature warming is the polar regions in the Northern Hemisphere (Fig. 1). The greater warming in the Arctic¹, compared with the global mean, is associated with a reduction in sea ice² and dynamical and radiative feedbacks³, and is widely attributed to anthropogenic climate change. But the fact that the warming is not spatially uniform raises the question of whether natural climate variability has a role in

driving it and causing regional climate change. On page 209 of this issue, Ding *et al.*⁴ show that the most prominent Arctic warming has occurred in northeastern Canada and Greenland, and that cooling in the tropical Pacific Ocean forced half of the warming in these two regions. The findings indicate that a substantial part of regional Arctic climate change is therefore a result of natural climate variability.

Much of the interannual and decadal variability in atmospheric climate can be described by the evolution of the leading modes of climate variability, such as the North Atlantic Oscillation (NAO). The NAO consists

of variations in the difference of sea-level atmospheric pressure between the Icelandic low-pressure system (Icelandic low) and the Azores high-pressure system (Azores high), and is most pronounced during boreal winter. In the positive phase of the NAO, there is a considerable difference in pressure between these two systems, with both the Icelandic low and the Azores high being intensified. In the negative phase, the two pressure zones are weakened and the difference between them is less.

The NAO is linked to changes in the intensity and location of the North Atlantic jet stream and storm track, and to large-scale temperature and precipitation variations over Europe, Greenland and North America. It is an intrinsic atmospheric phenomenon, but fluctuations in sea surface temperature (SST) can also affect it. In fact, it can be influenced by both variations in local North Atlantic SST and remote SST in the tropics⁵. Changes in tropical SSTs lead to changes in convection throughout the lowest portion of the atmosphere (the troposphere) at low latitudes, which in turn excite large-scale atmospheric waves called Rossby waves. These waves can propagate to mid- and high latitudes and affect the NAO.

In their study, Ding *et al.* demonstrate that Rossby waves and the NAO are involved in regional Arctic warming. Their finding that the Arctic warming in northeastern Canada and Greenland since 1979 is strongly driven by cooling in the tropical Pacific is supported by observational data indicating that warming in these two regions is not limited to the surface but also extends to the upper troposphere. The authors argue that it is unlikely that decadal temperature changes in the upper Arctic troposphere are locally forced by variations in surface temperature. They suggest instead that warming at the surface and in the troposphere are the result of atmospheric-circulation changes in the high troposphere, and that these changes are remotely forced. Specifically, Ding *et al.* show that the recent warming in northeastern Canada and Greenland is associated with a negative NAO phase driven by Rossby-wave activity caused by SST cooling in the tropical Pacific.

These results are confirmed by modelling experiments. The authors demonstrate that an atmospheric general circulation model forced by the observed tropical SST can simulate the connection between tropical Pacific SST cooling and regional Arctic tropospheric warming. However, they also show that coupled ocean-atmosphere climate models used in the fifth assessment report of the Intergovernmental Panel on Climate Change fail to reproduce the observed regional pattern of

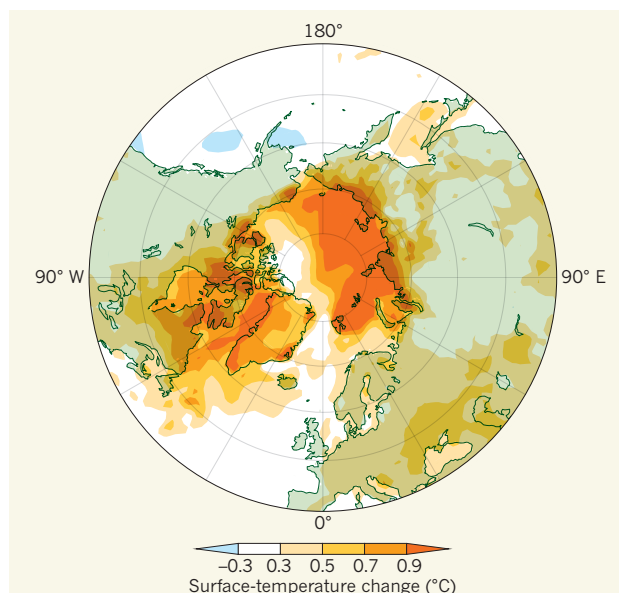


Figure 1 | Trend in annual mean surface temperature. The graphic shows the observed change per decade of annual mean surface and near-surface temperature for the period 1979–2012, based on the ERA-interim climate data set. The most marked warming has occurred in northeastern Canada, Greenland and north Siberia. (Adapted from Extended Data Fig. 1 of ref. 4.)

Arctic warming. Two plausible reasons for this failure are worth mentioning. First, the recent cooling in the tropical Pacific can probably be attributed to intrinsic variability of the climate system⁶, because it is not simulated by coupled climate simulations that incorporate observed changes in the concentration of greenhouse gases and aerosols. Second, despite continued improvements to climate models, it is still a challenge to simulate the influence of remote

climatic phenomena correctly.

By linking cooling in the tropical Pacific with trends in atmospheric circulation and regional Arctic warming, Ding and colleagues highlight the complexity of processes involved in regional climate change. Even remote climatic fluctuations can have a substantial impact. Improving the representation of such teleconnections in climate models should therefore remain a high priority for climate scientists. The authors also note the importance of natural internal climate variability for present and near-future regional Arctic climate. But as greenhouse-gas concentrations are likely to increase in the future, it is only a question of time before external forcing dominates regional Arctic warming. ■

Jürgen Bader is at the Max Planck Institute for Meteorology, 20146 Hamburg, Germany, and at Uni Climate, Uni Research & the Bjerknes Centre for Climate Research, Bergen, Norway.
e-mail: juergen.bader@mpimet.mpg.de

1. Graversen, R. G., Mauritsen, T., Tjernström, M., Källén, E. & Svensson, G. *Nature* **451**, 53–56 (2008).
2. Screen, J. A. & Simmonds, I. *Nature* **464**, 1334–1337 (2010).
3. Shindell, D. & Faluvegi, G. *Nature Geosci.* **2**, 294–300 (2009).
4. Ding, Q. *et al.* *Nature* **509**, 209–212 (2014).
5. Hoerling, M. P., Hurrell, J. W. & Xu, T. *Science* **292**, 90–92 (2001).
6. Kosaka, Y. & Xie, S.-P. *Nature* **501**, 403–407 (2013).

SYNTHETIC BIOLOGY

Construction of a yeast chromosome

One aim of synthetic biology is to generate complex synthetic organisms. Now, a stage in this process has been achieved in yeast cells — an entire yeast chromosome has been converted to a synthetic sequence in a stepwise manner.

DANIEL G. GIBSON & J. CRAIG VENTER

A biological cell is much like a computer — the genome can be thought of as the software that encodes the cell's instructions, and the cellular machinery as the hardware that interprets and runs the software. Advances in DNA technology have made it possible for scientists to act as biological 'software engineers', programming new biological 'operating systems' into

cells. Indeed, in 2010, the entire genome of the bacterium *Mycoplasma mycoides* was replaced with a rewritten synthetic genome, generating the first synthetic cell¹. Now, in a paper published in *Science*, Annaluru *et al.*² describe how they have begun rewriting the genome of a more complex organism, that of the yeast species *Saccharomyces cerevisiae*. The researchers report the design and generation of a functional, synthetic chromosome in this yeast, a milestone that they have

had in their sights for almost ten years.

A quote by the theoretical physicist Richard Feynman, “What I cannot create, I do not understand”, has inspired synthetic biologists around the world. But this does not necessarily mean that what can be created is understood. There is an enormous gap between our ability to build DNA and our understanding of the instructions it encodes. The production of a synthetic chromosome in yeast (the first such achievement in a eukaryote — the class of organisms comprising plants, animals and fungi) represents a step towards closing that gap. Generation of synthetic versions of widely used model organisms such as *S. cerevisiae* will enable scientists to investigate the requirements of life in eukaryotes, because synthetic organisms can be easily manipulated.

The process began at the computer, where Annaluru and colleagues downloaded the publicly available DNA sequence of a *S. cerevisiae* chromosome (chromosome III). Next, they designed genetic changes, which can be thought of as software edits, with the aim of introducing specific alterations into the chromosome. These edits included the deletion of dispensable DNA sequences; the incorporation of unique sequences to enable the researchers to differentiate between natural and engineered DNA; and the replacement of one particular region of DNA that terminates gene transcription with another that performs the same task.

In addition, the authors flanked each non-essential gene with DNA sequences designed to cause deletion of the flanked genes on a given signal. This flanking allows the size of the chromosome to be reduced, a feature designed to help determine the smallest cohorts of genes required to perform a given function or necessary for survival under a particular growth condition. This information is crucial if we are to write biological software in a more predictable fashion, thus ensuring that cells can be engineered to reliably carry out the tasks they have been programmed to perform.

To generate the synthetic chromosome, Annaluru and co-workers broke down their designer DNA sequence into overlapping stretches of 70 nucleotides, which were chemically synthesized in parallel. Students in the Build-a-Genome course at Johns Hopkins University in Baltimore, Maryland, stitched together these DNA stretches into constructs approximately 3 kilobases long, using established DNA-assembly methods^{3–6}.

Next, the original chromosome sequence was systematically replaced with the synthetic DNA *in vivo*, by introducing up to 12 overlapping 3-kb synthetic fragments at a time into the yeast cell, in 11 successive rounds of integration.

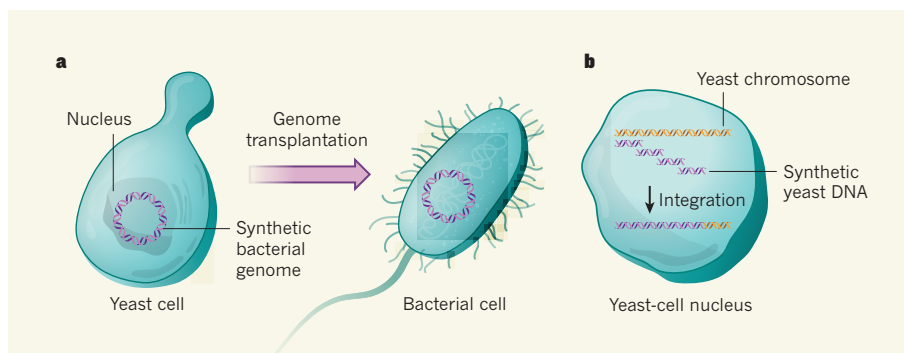


Figure 1 | Building synthetic genomes. **a**, To generate a synthetic bacterial cell, the entire synthetic bacterial genome is assembled in a yeast cell. In this setting, the bacterial genome is inactive because the yeast cell lacks the proteins required to turn on bacterial genes. After synthesis is complete, the synthetic genome is transferred to a compatible bacterial host cell, where it is activated and produces a synthetic cell. **b**, By contrast, Annaluru *et al.*² have generated a synthetic yeast chromosome within the yeast cell itself. To do this, they replaced the natural chromosome with chunks of synthetic DNA in a stepwise manner.

The product of this work is a yeast strain that contains an extensively engineered chromosome III, and that grows just as well as the original strain. The 273-kb designer chromosome contains more than 50,000 sequence alterations, and is 14% shorter than the natural sequence.

This work is important because it begins to address unanswered questions about how genome design can be used to manipulate the rules of biology in a eukaryotic model. For example, can superfluous DNA between genes be removed? Can unnecessary genetic code be altered? So far, the sequence alterations made by Annaluru and colleagues have not reduced the fitness of the yeast, which bodes well for future modifications.

If two or more genes perform a similar function, can one be deleted? Which combinations of the 5,000 yeast genes that are thought to be dispensable can be simultaneously removed? The authors have already begun cataloguing genes that can be simultaneously deleted without adversely affecting the fitness of the yeast. Answering these fundamental questions is a goal that the synthetic-yeast group shares with efforts to recode the genome of the bacterium *Escherichia coli*⁷, and with current research programmes^{1,8} that aim to understand better how to build bacterial genomes containing only the genes necessary to sustain life.

Two bacterial genomes have previously been chemically synthesized — the 583-kb genome of *Mycoplasma genitalium*⁴ and the 1,078-kb *M. mycoides* genome¹ — and so synthesis of a chromosome of 273 kb is not in itself unusual. The genomes of both *M. genitalium* and *M. mycoides* were assembled and propagated in yeast, acting as an extra chromosome. Like incompatible Macintosh software in a Windows computer, synthetic bacterial genomes are unable to ‘boot up’ in yeast-cell machinery, and so cannot produce self-replicating bacterial cells in this setting. In a process called genome transplantation^{9,10}, the complete bacterial genome must be moved to a bacterial host with compatible

hardware, converting the recipient cell into a new, synthetic species (Fig. 1a). By contrast, Annaluru *et al.* gradually converted the natural yeast chromosome into a fully functional designer chromosome within their target species, the yeast cell itself (Fig. 1b).

In comparison to the completely synthetic *M. mycoides* cell¹, chromosome III accounts for less than 3% of the yeast genome. Now the question is: can these design rules be successfully applied across the entire yeast genome? The scientists attempting to generate synthetic yeast cells still have a long way to go before they make a fully reprogrammed yeast genome. However, by demonstrating the success of their design principles and assembling an international team of scientists to build the remaining 15 chromosomes, Annaluru and colleagues have laid the groundwork for making this happen in the near future.

Advances such as these are of interest to the entire field of synthetic biology. Each innovation not only enhances our general understanding of biology, but also creates a framework on which we can build and expand, allowing us to work towards the goal of basing our economy on synthetic biology — a development that will have a positive impact on all of society. ■

Daniel G. Gibson and J. Craig Venter
are at the J. Craig Venter Institute, La Jolla, California 92037, USA, and at Synthetic Genomics, La Jolla.
e-mail: dgibson@jcv.org

- Gibson, D. G. *et al.* *Science* **329**, 52–56 (2010).
- Annaluru, N. *et al.* *Science* **344**, 55–58 (2014).
- Gibson, D. G. *et al.* *Proc. Natl Acad. Sci. USA* **105**, 20404–20409 (2008).
- Gibson, D. G. *et al.* *Science* **319**, 1215–1220 (2008).
- Gibson, D. G. *et al.* *Nature Methods* **6**, 343–345 (2009).
- Annaluru, N. *et al.* *Methods Mol. Biol.* **852**, 77–95 (2012).
- Isaacs, F. J. *et al.* *Science* **333**, 348–353 (2011).
- Glass, J. I., Hutchison, C. A. III, Smith, H. O. & Venter, J. C. *Mol. Syst. Biol.* **5**, 330 (2009).
- Lartigue, C. *et al.* *Science* **325**, 1693–1696 (2009).
- Lartigue, C. *et al.* *Science* **317**, 632–638 (2007).



COSMOLOGY

A virtual Universe

A numerical simulation of cosmic structure formation reproduces both large- and smaller-scale features of a representative volume of the Universe from early in its history to the present day. [SEE ARTICLE P.177](#)

MICHAEL BOYLAN-KOLCHIN

Perhaps the greatest triumph of modern cosmology is that a model with only six parameters can explain the vast majority of observational data from the first minutes of the Universe to the present day¹. This standard model posits that 95% of the Universe today is composed of enigmatic 'dark matter' and 'dark energy'. Paradoxically, modelling the dynamics of the remaining 5% — normal, 'baryonic' matter — has proved to be the more challenging task. On page 177 of this issue, Vogelsberger *et al.*² describe a numerical simulation of the formation of cosmic structure that captures both the large-scale distribution of baryonic material and its properties in individual galactic systems through cosmic time.

Tracking the evolution of baryonic matter is a daunting undertaking because of the huge range of physical scales involved in the processes that shape galaxies and larger structures (Fig. 1). To cover a representative portion of the Universe, cosmologists must study cosmic volumes that are at least 100 million parsecs (326 million light years) across. By contrast, the natural scale of star formation is approximately 1 parsec, and accretion of gas by supermassive black holes occurs on even smaller scales. Numerical simulations have long been the tool of choice for tackling these problems. But even with the most powerful supercomputers, it has been impossible to run a simulation large enough to model the large-scale

distribution of gas, stars and dark matter while keeping sufficient detail to accurately capture individual galaxies. It turns out that simulating the Universe can be a difficult endeavour.

Vogelsberger and colleagues attack this problem from all sides. Their simulation — named Illustris — incorporates more than 10 billion individual cells to represent the gas in the simulation volume; this is nearly a tenfold increase on its predecessors. The numerical code³ used to perform their simulation employs a new approach, based on an unstructured and adaptive computational mesh that follows fluid flows, to solve the equations that describe the time evolution of baryonic matter within cosmic structures. And finally, the physical phenomena included in the simulation are rich and complex: the authors consider gas cooling, stellar evolution, energy input from supernova explosions, production of chemical elements, and gas accretion by supermassive black holes (along with accompanying radiative feedback), to list just a few.

If this all sounds somewhat complicated, do not be fooled: it is extremely complicated. Many of these processes are not understood from first principles, and they interact in complex, nonlinear ways. Additionally, the relevant physical scales are often (much) smaller than can be directly resolved, even with Illustris. This requires computationally efficient models that accurately encapsulate the underlying physics. Running the simulation was therefore no mean feat: it took approximately 16 million

CPU (central processing unit) hours. The end result, however, is a simulated Universe that looks an awful lot like the real one.

A mock observation of Illustris set to mimic the Hubble Ultra Deep Field⁴, the deepest picture of the cosmos ever taken, can easily pass for the real thing when the two are viewed side by side (see Fig. 1b, c of the paper²). Images of galaxies from the simulation are also impressively realistic (see Fig. 1a of the paper), an accomplishment that has previously been possible only for simulations of individual galaxies. This is not just window dressing: such procedures allow direct and meaningful confrontation of theory with data. A wide array of quantitative measures agrees with observations of the real Universe as well. For example, previous generations of simulations had great difficulty capturing the observed distribution of elements heavier than hydrogen and helium contained in stars. Illustris reproduces these observations, not just for the Universe as a whole, but also as a function of galactic stellar mass. In addition, the simulation matches the abundance of these heavier elements in dense gas clouds.

Of course, Illustris does not mark the end of cosmological simulations of galaxy formation. Although its computational volume is immense, it is not large enough to model the formation of rare cosmological objects (for example, powerful black holes observed in the early Universe). And the level of detail is still not fine enough to study the faintest galaxies surrounding the Milky Way. Star formation in low-mass galaxies occurs earlier and faster in Illustris than in the real Universe, a difficulty shared by almost all models of galaxy formation⁵. However, such issues point the way for future advances.

One clear goal for observers and theorists alike is to understand in detail the ways in which energy and momentum from evolving and exploding stars affect the properties of gas in and around galaxies^{6,7}. A promising computational approach is to combine large-volume

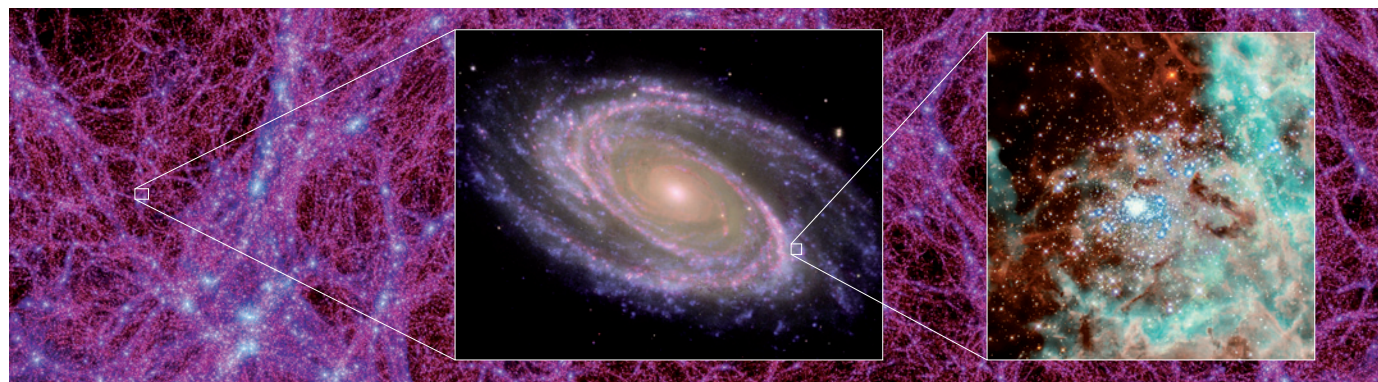


Figure 1 | The multi-scale nature of galaxy formation. Studying the formation of galaxies such as the spiral galaxy Messier 81 (left inset) requires simulating their large-scale environments (background image) as well as understanding the birthplaces of individual stars (right inset). The scale of the background image is approximately 150 million parsecs, and each inset image zooms in by a factor of approximately 10,000. It is this huge range in scales that

makes *ab initio* modelling of galaxy formation such a challenge. Vogelsberger *et al.*² were able to simulate a representative volume of the Universe while still resolving details for individual galaxies. (Background image: M. Boylan-Kolchin *et al.*/Max Planck Inst. Astrophys.; left inset: NASA/JPL-Caltech/ESA/Harvard-Smithsonian CfA; right inset: NASA, N. Walborn & J. Maíz-Apellániz (STSI), R. Barbá (La Plata Observatory).)

simulations such as Illustris, which are necessarily limited in the level of detail they can resolve, with simulations that focus all of their power on individual galaxies, which sacrifice the ability to study galaxy formation in a statistical manner. If the knowledge gained from these focused simulations can be incorporated into large-scale computational efforts, ever more accurate numerical realizations of the Universe should be possible, guiding our investigations into the underlying physical processes.

Computationally, the ability to reach the

scales necessary to model star formation directly, while encompassing thousands of galaxies similar to the Milky Way, is still a distant dream. But thanks to the Illustris simulation, creating a realistic virtual Universe of gas, stars, black holes and dark matter is already a reality. ■

Michael Boylan-Kolchin is in the Department of Astronomy, University of Maryland, College Park, Maryland 20742-2421, USA.
e-mail: mbk@astro.umd.edu

1. Planck Collaboration. *Astron. Astrophys.* (in the press); preprint at <http://arxiv.org/abs/1303.5076> (2014).
2. Vogelsberger, M. *et al. Nature* **509**, 177–182 (2014).
3. Springel, V. *Mon. Not. R. Astron. Soc.* **401**, 791–851 (2010).
4. Illingworth, G. D. *et al. Astrophys. J. Suppl. Ser.* **209**, 6 (2013).
5. Weinmann, S. M. *et al. Mon. Not. R. Astron. Soc.* **426**, 2797–2812 (2012).
6. Agertz, O., Kravtsov, A. V., Leitner, S. N. & Gnedin, N. Y. *Astrophys. J.* **770**, 25 (2013).
7. Hopkins, P. F. *et al.* Preprint at <http://arxiv.org/abs/1311.2073> (2013).

BIODIVERSITY

Supply and demand

Data from Himalayan songbirds suggest that the rate-limiting step in biodiversity production may not be the speed of speciation, but rather the speed at which new niches are created. [SEE LETTER P.222](#)

ARNE O. MOOERS

There are currently about 10,000 named bird species on the planet, but only half as many mammals. What dictates the number of species in any given group? In this issue, Price *et al.*¹ (page 222) marry the ecologies of bird species living in close proximity in northern India with their ages to argue that it is not the rate at which new species form but the rate at which new niches are created that limits how many species actually persist. By building a new evolutionary tree linking all 358 songbird species inhabiting a Himalayan slope, the authors find that it has taken an average of 7 million years for related species to end up together on the landscape.

This seems a long time: bird speciation generally involves geographic separation and, from beginning to end, it usually takes less than 3 million years for species-level mutual disdain to evolve; new species can populate adjacent landscapes and join their close relatives in the blink of a geological era. That said, species that occupy the same niche rarely occur together in nature. So Price and colleagues argue that the roughly 4 million extra years it seems to take for species to establish themselves alongside their relatives means they have had to wait that long for available niches to arise. This suggests that finding a new role in nature may be the limiting step in diversification.

Given the ambiguous signals that evolutionary trees have offered us so far, using a partial evolutionary tree linking coexisting species to make an argument about large-scale processes is audacious. Previous studies that have compared the lengths of recent and older branches on evolutionary trees to measure diversification² generally have found that more-recent branches are too long relative to

expectations from simple evolutionary models. Such data often suggest both that extinction is non-existent (which, of course, is not true) and that diversification has slowed towards the present³.

There are at least three reasons for this odd pattern of slowing diversification, and Price *et al.* consider them all. Mundanely, inferring trees is hard, and recent branches on evolutionary trees may be biased to be too long relative to older branches⁴. Luckily, the authors could compare their tree to a recent and independently derived one that includes nearly all the same species⁵. Crucially, although the two trees differ greatly in total age and shape, they

agree almost exactly on the 7-million-year time lag for the coexistence of relatives.

More interestingly, an apparent slowdown in diversification is expected because speciation takes time: we miss more-recent diversification events because we do not recognize separated populations that will eventually become new species as new species⁶. This means that recent speciation events have not yet been recorded, so recent branches will be longer than they should be. Many evolutionary trees are well described by such a process⁷. However, Price and colleagues sidestep this issue with a particularly nimble piece of logic. Because most speciation takes place when populations are separated geographically, and given that lineages that occupy the same niche rarely coexist, there can be few or no unrecognized species in a sample of species living in proximity. Nascent species may exist elsewhere, but the recent branches of these authors' tree are certainly not too long because of a species-recognition problem.

Most intriguing is to consider the possibility that slowdown is a common fate in diversification. Following the invasion of relatively empty

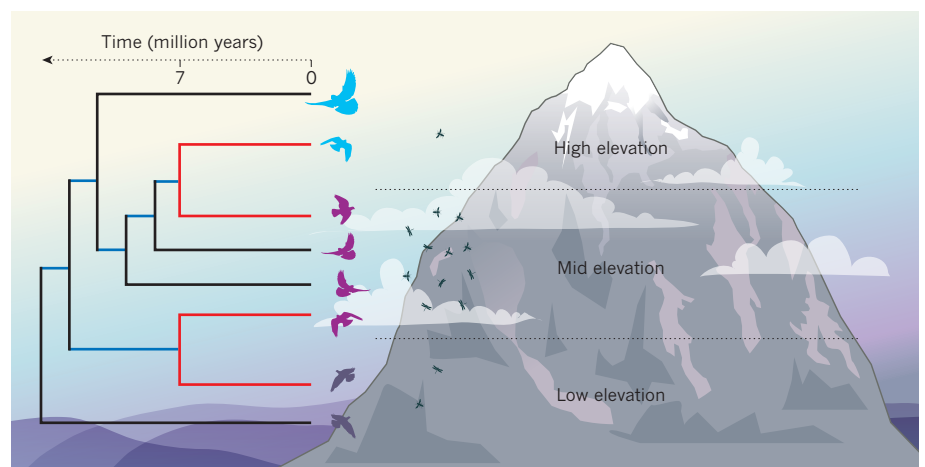


Figure 1 | Himalayan songbird assemblage. Although bird species generally take less than 3 million years to form, Price *et al.*¹ show that closely related species (those joined by the recent (red) branches of the evolutionary tree) that are found together in the landscape are separated by about 7 million years — these recent branches are much longer than most of the deeper (blue) branches. Their tree also reveals that closely related bird species tend to occupy different elevations; there are more species at mid-elevations where there is more food; and variation in songbird size and shape is highest at high elevations, which are species-poor and therefore competition-poor. These patterns are all consistent with a model in which the ultimate brake on diversification is ecological opportunity.

niche space — by dispersal to a new region or by the evolution of new ways to make a living — an initial flurry of speciation fills up an initial set of niches, but then new species become established only when others go extinct or new niches are created⁸.

Such an ecological-demand model (in contrast to evolutionary supply) also fits many evolutionary trees well³, and Price and colleagues' data on songbirds tally with this story (Fig. 1). Other data from the region suggest that recent ecological opportunity may be linked to changes in elevation, and Price *et al.* indeed find that close relatives differ most often in the elevation at which they live. The well-known pattern of highest species richness at mid-elevations⁹ is, in their data, associated with more food (and so more niches) rather than with more speciation. And songbirds span a greater range of sizes and shapes at species-poor high elevations, perhaps owing to release from competition with other groups of birds that could not establish themselves.

There is little question that an ecological-demand model of biodiversity offers up a grand narrative: nascent species arise in ample supply as isolated populations, but these expand their ranges and fill landscapes only if, and only when, those landscapes have ecological room for them. Many new species might therefore wither on the vine and never reach this expansion phase. This model implies a decoupling of the rate of speciation from the rate of subsequent establishment, a novel pattern that has recently been documented for flies and birds¹⁰.

An ecological-demand model also raises several questions. Evolutionary biologists have long appreciated that close relatives compete strongly¹¹. However, we have only recently re-appreciated that speciation is often intimately linked to diverging ecologies¹². Therefore, to evaluate the ecological-demand model, we need to know the relative contributions of ecological¹³ and non-ecological speciation¹⁴ to biodiversity production. And, given that the model suggests that the number of species in a group is related to niche availability¹⁰, we need to understand both why there are more niches for some groups (say, birds) than for others (say, mammals), and whether it is really true that, at any one time, many landscapes are full to the brim with biodiversity. ■

Arne O. Mooers is in the Biology Department and the Human Evolutionary Studies Program, Simon Fraser University, Burnaby, British Columbia V5A1S6, Canada. e-mail: amooers@sfu.ca

1. Price, T. D. *et al.* *Nature* **509**, 222–225 (2014).
2. Nee, S., Mooers, A. O. & Harvey, P. H. *Proc. Natl Acad. Sci. USA* **89**, 8322–8326 (1992).
3. Morlon, H., Potts, M. D. & Plotkin, J. B. *PLoS Biol.* **8**, e1000493 (2010).
4. Revell, L. J., Harmon, L. J. & Glor, R. E. *Syst. Biol.* **54**, 973–983 (2005).

5. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. *Nature* **491**, 444–448 (2012).
6. Purvis, A., Orme, C. D. L., Toomey, M. H. & Pearson, P. N. in *Speciation and Patterns of Diversity* (eds Butlin, R., Schluter, D. & Bridle, J.) 278–300 (Cambridge Univ. Press, 2009).
7. Etienne, R. S. & Rosindell, J. *Syst. Biol.* **61**, 204–213 (2012).
8. Rabosky, D. L. & Lovette, I. J. *Proc. R. Soc. B* **275**, 2363–2371 (2008).
9. Rahbek, C. *Ecography* **18**, 200–205 (1995).

10. Rabosky, D. L. & Matute, D. R. *Proc. Natl Acad. Sci. USA* **110**, 15354–15359 (2013).
11. Darwin, C. *On the Origin of Species* (Murray, 1859).
12. Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
13. Nosil, P. *Ecological Speciation* (Oxford Univ. Press, 2012).
14. Svenson, E. I. *Organisms Diversity Evol.* **12**, 229–240 (2012).

This article was published online on 30 April 2014.

ECOLOGY

Diversity in the afterlife

Field experiments that varied the composition of both plant litter and the organisms that break it down have revealed that, across ecosystems, lower biodiversity slows the rate of litter decomposition. [SEE LETTER P.218](#)

JENNIE R. MCLAREN

Despite our increased awareness of biodiversity loss, and attempts to respond to it, the global rate of species extinction does not seem to be slowing¹. Concern about how these losses may affect the way that ecosystems function has led to decades of experiments looking for connections between the properties of ecosystems and the numbers and types of species that live in them. Using field experiments and meta-analyses, ecologists have determined that decreased diversity reduces the efficiency of resource capture by communities and the conversion of those resources to biomass². Most biomass created by plants is not consumed by herbivores, but becomes 'litter' after a plant dies. Do the effects of diversity die with the plants? On page 218

of this issue, Handa *et al.*³ suggest that the answer is no, by showing that both the diversity of dead plants and the diversity of organisms that decompose them contribute to determining how quickly this material is recycled back into ecosystems.

Although experiments and meta-analyses have given us confidence in our predictions of the effects of mixing together living plants, experiments in which different litter types are mixed rarely show predictable results^{4–6}. Meta-analyses of these studies have been limited by different experimental designs, site-specific effects and the potential influence of the type rather than the number of species. Handa and collaborators attacked this problem using a series of coordinated field experiments, at sites ranging from the sub-Arctic to the tropics, in which they manipulated the diversity of both

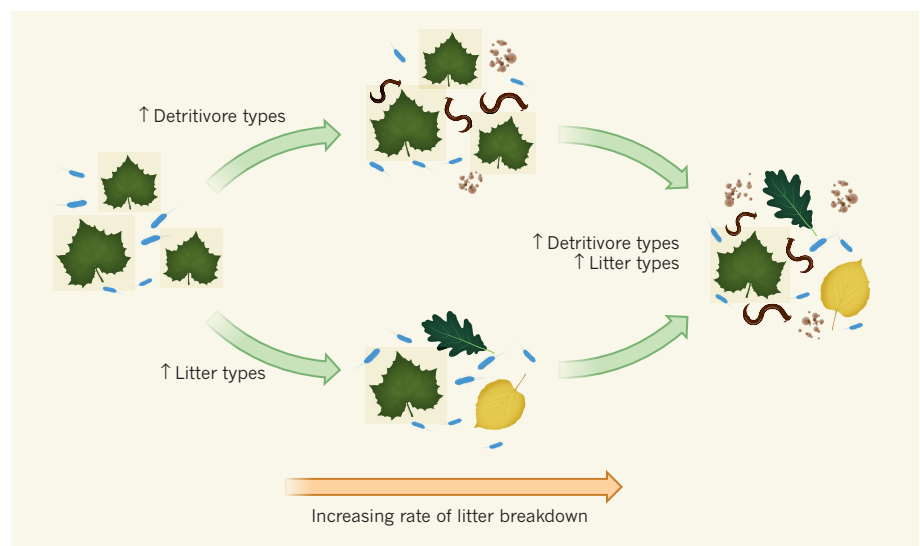


Figure 1 | Composition and decomposition. Handa *et al.*³ show that increasing the diversity of either plant litter or detritivores — invertebrates and microorganisms that break down litter — independently increases rates of litter breakdown. Simultaneously increasing both litter and detritivore diversity results in the largest increases in decomposition rate.

plants and detritivores — the invertebrates and microorganisms that break down litter. They found that a reduction in the diversity of either slows the rate at which litter is decomposed, regardless of the location of the experiment (Fig. 1).

Decomposition rates are an important determinant of the global carbon budget, affecting not only the incorporation of vegetation carbon into soils, but also the early stages of its release back to the atmosphere. That Handa *et al.* found effects of diversity on this process across biomes and ecosystems suggests a remarkable consistency in the consequences of biodiversity loss.

The authors manipulated detritivore diversity by varying the types of container the litter was decomposed in, and thus the organisms that could access and break down the litter. They show that the diversity of these organisms had even more influence than the composition of the litter community itself — the more types of detritivore, the faster the litter was decomposed. This finding provides experimental support for the results of a meta-analysis⁵ that concluded that the effects of consumer diversity on decomposition are generally more important than the effects of resource diversity.

Handa and colleagues' study is also one of the first to provide evidence for a potential mechanism driving the effects of higher litter diversity: the movement of nitrogen between different litter types. Laboratory studies had previously shown that such nitrogen movement was possible⁷, but whether it could occur in natural settings, and whether it would lead to changes in decomposition, was unknown until now.

Although nitrogen movement between leaves had been predicted to follow a simple gradient — from leaves with more nitrogen to leaves with less — it may instead be driven by carbon quality. If microbes grow quickly on litter from plants with high-quality carbon, this would increase the microbes' need for nitrogen, and when it is not available locally they may access it from the neighbouring litter. Handa *et al.* found evidence for nitrogen movement between litter types across biomes and ecosystems, and always between the same two litter types (from nitrogen-fixing litter species to rapidly decomposing ones). This is an exciting result that will increase our ability to predict the effects of biodiversity loss on carbon cycling and storage.

However, there is still no direct evidence for nitrogen translocation — in the latest experiments, it was inferred from patterns in nutrient loss or gain as the litter decomposed. The researchers report that when the nitrogen-fixing and fast-decomposing litter species are decomposed together, the amount of nitrogen present in one goes up while the amount in the other goes down, which is best explained by nitrogen moving between species. But it is possible that the nitrogen

moving into the litter is coming from elsewhere, including the surrounding soil. Given that the predictive power of Handa and colleagues' results hinges on knowledge about underlying mechanisms, a key next step will be to confirm the implied mechanism with direct tests in the field. One possibility would be to perform experiments using litter with radiolabelled nitrogen in the field, which have so far been done only in the lab.

We are increasingly aware that understanding the impacts of biodiversity loss requires recognizing that not all species are at equal risk of extinction⁸. The identity of species lost has been shown to have large effects on plant productivity⁹ and decomposition³, which results in a range of potential outcomes for different extinction scenarios. Thus, being able to accurately predict the effects of species loss at a broader scale will require both generalizable rules linking certain types of species to the effects of their loss, and knowledge of the mechanisms that cause those effects.

Handa and colleagues' study brings us closer to that goal by providing three things: general patterns linking litter and detritivore diversity

to decomposition; evidence for a mechanism driving that link; and evidence that these patterns and mechanism are the same regardless of whether they occur in an Arctic stream or a forest in the tropics. With this information in hand, we can strive towards linking patterns to realistic extinction scenarios, and thus predict probable outcomes of biodiversity loss. ■

Jennie R. McLaren is in the Department of Biological Sciences, University of Texas at El Paso, El Paso, Texas 79968, USA.
e-mail: jrmclaren@utep.edu

1. Butchart, S. H. *et al.* *Science* **328**, 1164–1168 (2010).
2. Cardinale, B. J. *et al.* *Am. J. Bot.* **98**, 572–592 (2011).
3. Handa, I. T. *et al.* *Nature* **509**, 218–221 (2014).
4. Hättenschwiler, S., Tiunov, A. V. & Scheu, S. *Annu. Rev. Ecol. Evol. Syst.* **36**, 191–218 (2005).
5. Srivastava, D. S. *et al.* *Ecology* **90**, 1073–1083 (2009).
6. Cardinale, B. J. *et al.* *Nature* **486**, 59–67 (2012).
7. Schimel, J. P. & Hättenschwiler, S. *Soil Biol. Biochem.* **39**, 1428–1436 (2007).
8. Wardle, D. A., Bardgett, R. D., Callaway, R. M. & Van der Putten, W. H. *Science* **332**, 1273–1277 (2011).
9. Hooper, D. U. *et al.* *Nature* **486**, 105–108 (2012).

STRUCTURAL BIOLOGY

Action at a distance in a light receptor

A tour de force of X-ray scattering has yielded structures of a phytochrome photoreceptor in its dark and illuminated states, showing how localized protein refolding magnifies a light signal to form a cellular message. [SEE LETTER P.245](#)

ANNA W. BAKER & KATRINA T. FOREST

Sensor proteins allow organisms to perceive and appropriately respond to environmental changes. They are molecular machines comprised of sensory domains and effector modules; a specific signal is received by the former and communicated to the latter through structural changes, creating an output activity that directs a cellular response. Phytochromes are dimeric sensor proteins that specifically absorb red and near-infrared light using a covalently tethered chromophore molecule housed in a light-sensitive (photosensory) core. In this issue, Takala *et al.*¹ (page 245) demonstrate how visible light is interpreted and spatially magnified by the phytochrome dimer in a chain reaction that uses key features of the protein's three-dimensional structure.

Phytochromes regulate most of the responses of plants to light, including germination, flowering and shade avoidance, and are thus fundamental to agriculture². They are also

widespread in microbes³, and phytochromes found in cyanobacteria are thought to be ancestors of those in plants². The proteins signal by switching between two stable conformations, which correspond to the dark and illuminated states. Absorption of light by the chromophore alters a local network of hydrogen bonds and van der Waals interactions, and ultimately induces the illuminated conformation.

How these minute changes (of the order of ångströms) are transferred through the protein to regulate the distant effector domain is a long-standing puzzle. One proposed model, also applicable to blue-light receptors and sensors that react to chemical signals, involves the rotation of central helices connecting the sensory and effector modules⁴. In phytochromes, a structural element known as the tongue also spans the two modules and is likewise expected to be involved in signal transduction⁵.

The inherent dynamism of the dark and illuminated states confounds structural studies of phytochromes. Several dark-state crystal

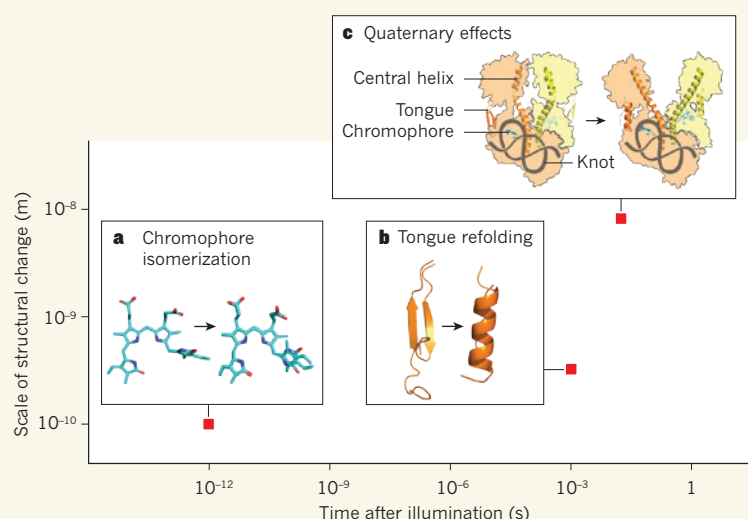


Figure 1 | Signal magnification in phytochrome light sensors. **a**, Light-induced isomerization of the chromophore in phytochromes causes atomic-scale structural changes within picoseconds of illumination. **b**, Takala *et al.*¹ report that these changes lead to rearrangement of a crucial secondary-structure element in the 'tongue' region of the phytochrome: a β -hairpin unravels and an α -helix forms. **c**, Within 30 milliseconds, the closed quaternary structure adopted in the dark state opens to a Y-shaped conformation in the illuminated state, with effector domains (structure unknown) at the distal tips of the Y. The phytochrome's characteristic figure-of-eight knot may rigidify the chromophore-binding domain, thereby restricting light-driven motion to the hinge in the central helix of each dimer subunit. The two subunits of the phytochrome are shown in orange and yellow; red squares indicate approximate times and length scales of the changes.

structures of domains of the photosensory core have been solved^{6,7}, but finding a way to consistently trap the phytochrome in the illuminated state for crystallography has been a major obstacle. Takala and colleagues overcame this problem by taking advantage of the fact that the phytochrome from the bacterium *Deinococcus radiodurans* relaxes slowly *in vitro* from its illuminated state to its dark conformation. The authors therefore delivered periodic red-light pulses to growing crystals of the protein, to enrich the population of molecules in the illuminated state, and thereby enable a stable crystal of the sensor in that conformation to be grown. This allowed low-resolution X-ray structures to be solved for the complete photosensory core in the dark and illuminated crystalline states.

Next, the researchers performed molecular dynamics simulations based on the crystal structures to generate structures of the protein in solution. These were validated by comparison with X-ray-scattering data for the photosensory core in solution. Together, the crystal and solution structures support a model in which chromophore isomerization resulting from red-light absorption ruptures adjacent hydrogen bonds, causing a β -hairpin structural feature in the tongue to unfold and nearby atoms in the same malleable structural unit to wind into an α -helix. The tongue thus contracts by 2.5 Å to relieve a sharp (approximately 50°) bend at a hinge in the central helix. The resulting illuminated state has a distinct Y shape, in which the arms of the Y are as much as 8 nanometres apart. In an experimental

tour de force, the authors combined time-resolved X-ray-scattering data collected over several timescales in solution, and found that the phytochrome magnifies light-induced atomic-scale changes (produced within picoseconds⁶; 1 picosecond is 10^{-12} s) to the nanometre scale in milliseconds (Fig. 1).

Models of signal transduction in which the central helix of the phytochrome rotates are not supported by these findings, but are supplanted by a scenario in which the helix straightens in response to refolding of the tongue when illuminated. This explains why two tongue conformations have previously been seen for phytochromes^{6,7}. The new model is supported by earlier observations⁸ of light-induced effects near the chromophore, and plausibly connects structural changes from the atomic to the nanometre scale. A central tenet of the mechanism is that the secondary structure of the tongue changes, but could it be that the observed conformations occur only in crystals, and not in solution? Such concerns are largely allayed by the triangulation of results obtained using molecular dynamics, solution scattering and protein crystallography.

The tongue-refolding model requires that dramatic movements within the protein are manifested through a plastic region in the central helix; for this to occur, kinetic energy must not be misdirected into random motion throughout the photosensory core. Phytochromes contain an unusual figure-of-eight knot⁹, and we propose that this limits the number of degrees of conformational freedom available to the protein when an energetic red-light

photon is absorbed. This rigidity presumably funnels the protein's response to the appropriate changes in secondary structure along the chromophore–tongue–effector-domain trajectory so that signal transduction can proceed.

Takala and co-workers provide a picture of phytochrome signal transduction that suggests a mechanism for how the activity of the effector domain is regulated by red light. In *D. radiodurans*, the phytochrome effector module is a histidine kinase. In the illuminated state, the ends of the two Y arms each bear one of these enzymatic modules. We infer that the trajectory of the central helix observed in each monomer of the illuminated structures continues through the kinase domain, although this remains to be proven experimentally. In any case, strain induced by straightening of the long central helix will force the effector modules into a new orientation relative to one another.

Histidine kinases act by transferring a phosphate group to their substrate across a subunit–subunit interface¹⁰. In the case of the phytochrome, it is logical to conclude that enzymatic activity occurs in the dark, when the active site of an effector module is close to the phosphorylation site of its partner, but halts in the light, when the modules are separated¹¹. Unfortunately, crystallization of dynamic histidine kinases suffers from similar roadblocks to those associated with phytochrome crystallization. A complete phytochrome structure at near-atomic resolution in both states therefore remains the ultimate goal in the field, along with accompanying enzymological and physiological studies to support any structural model that shows how local structural changes are magnified into global ones. ■

Anna W. Baker and Katrina T. Forest are in the Department of Bacteriology, University of Wisconsin–Madison, Madison, Wisconsin 53706, USA.
e-mails: forest@bact.wisc.edu;
awbaker@wisc.edu

1. Takala, H. *et al.* *Nature* **509**, 245–248 (2014).
2. Quail, P. H. *Nature Rev. Mol. Cell Biol.* **3**, 85–93 (2002).
3. Mandalari, C., Losi, A. & Gärtner, W. *Photochem. Photobiol. Sci.* **12**, 1144–1157 (2013).
4. Möglich, A. & Moffat, K. J. *Mol. Biol.* **373**, 112–126 (2007).
5. Anders, K., Daminelli-Widany, G., Mroginiski, M. A., von Stetten, D. & Essen, L.-O. *J. Biol. Chem.* **288**, 35714–35725 (2013).
6. Yang, X., Kuk, J. & Moffat, K. *Proc. Natl Acad. Sci. USA* **105**, 14715–14720 (2008).
7. Essen, L.-O., Mailliet, J. & Hughes, J. *Proc. Natl Acad. Sci. USA* **105**, 14709–14714 (2008).
8. Yang, X., Ren, Z., Kuk, J. & Moffat, K. *Nature* **479**, 428–432 (2011).
9. Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. *Nature* **438**, 325–331 (2005).
10. West, A. H. & Stock, A. M. *Trends Biochem. Sci.* **26**, 369–376 (2001).
11. Giraud, E. *et al.* *J. Biol. Chem.* **280**, 32389–32397 (2005).

This article was published online on 30 April 2014.

Properties of galaxies reproduced by a hydrodynamic simulation

M. Vogelsberger¹, S. Genel², V. Springel^{3,4}, P. Torrey², D. Sijacki⁵, D. Xu³, G. Snyder⁶, S. Bird⁷, D. Nelson² & L. Hernquist²

Previous simulations of the growth of cosmic structures have broadly reproduced the ‘cosmic web’ of galaxies that we see in the Universe, but failed to create a mixed population of elliptical and spiral galaxies, because of numerical inaccuracies and incomplete physical models. Moreover, they were unable to track the small-scale evolution of gas and stars to the present epoch within a representative portion of the Universe. Here we report a simulation that starts 12 million years after the Big Bang, and traces 13 billion years of cosmic evolution with 12 billion resolution elements in a cube of 106.5 megaparsecs a side. It yields a reasonable population of ellipticals and spirals, reproduces the observed distribution of galaxies in clusters and characteristics of hydrogen on large scales, and at the same time matches the ‘metal’ and hydrogen content of galaxies on small scales.

The initial conditions for structure formation in the Universe are tightly constrained from measurements of anisotropies in the cosmic microwave background radiation¹. However, previous attempts to reproduce the properties of the observed cosmological structures with computer models have shown only limited success. No single, self-consistent simulation of the Universe was able to simultaneously predict statistics on large scales, such as the distribution of neutral hydrogen or the galaxy population of massive galaxy clusters, together with galaxy properties on small scales, such as the morphology and detailed gas and stellar content of galaxies. The challenge lies in following the baryonic component of the Universe using hydrodynamic simulations^{2–4}, which are required to model gas, stars, supermassive black holes (SMBHs) and their related energetic feedback. The vast computational challenges of these simulations have forced previous attempts to focus either on simulating only small portions of the Universe or to employ a coarse resolution such that internal characteristics of galaxies could not be resolved. Furthermore, the large dynamic range of length scales in the problem forces common hydrodynamic solvers to sacrifice accuracy. Finally, poorly understood small-scale processes such as star formation and accretion onto SMBHs are coupled to galactic and super-galactic scales, introducing large uncertainties into modelling techniques.

Rapid advances in computing power combined with improved numerical algorithms and more faithful models of the relevant physics have allowed us to produce a simulation (named Illustris) that simultaneously follows the evolution of dark matter and baryons in detail. Starting approximately 12 million years (Myr) after the Big Bang, our simulation tracks the evolution of more than 12 billion resolution elements in a volume of $(106.5 \text{ Mpc})^3$ up to the current epoch (redshift $z = 0$). This allows us to achieve a dark-matter mass resolution of 6.26×10^6 times that of the Sun (M_\odot), and a baryonic mass resolution of $1.26 \times 10^6 M_\odot$. The smallest scale over which the hydrodynamics is resolved is 48 pc, whereas gravitational forces are resolved down to 710 pc at $z = 0$ and to even smaller scales at high redshifts (for example, 473 pc at $z = 2$). Our calculation therefore overcomes the problems of previous hydrodynamic simulations which either did not cover a large enough portion of the Universe to be representative, lacked adequate resolution, or failed to reach the present epoch. Apart from having a large

volume and improved resolution, our simulation is evolved with the novel hydrodynamic algorithm AREPO⁵, which uses a moving unstructured Voronoi tessellation in combination with a finite volume approach (Methods). Finally, we employ a numerically well-posed and reasonably complete model for galaxy formation physics, which includes the formation of both stars and SMBHs, and their effects on their environments in forms of galactic super-winds driven by star formation, as well as radio bubbles and radiation proximity effects caused by active galactic nuclei (AGNs; see Methods).

Unlike previous attempts, we find a mix of galaxy morphologies ranging from blue spiral galaxies to red ellipticals, with a hydrogen and ‘metal’ (that is, all elements other than hydrogen and helium) content in good agreement with observational data. At the same time, our model predicts correctly the large-scale distribution of neutral hydrogen, and the radial distribution of satellite galaxies within galaxy clusters. Our results therefore demonstrate that the Λ cold dark matter (Λ CDM) model can correctly describe the variety of observational data on small and large scales in our Universe. It also predicts a strong, scale-dependent impact of baryonic effects on the dark-matter distribution, at a level that has significant implications for future precision probes of cosmology.

Observing the model Universe

The simulation volume contains 41,416 galaxies at $z = 0$ that are resolved with more than 500 stellar resolution elements. Our model yields a population of non-star-forming elliptical galaxies, star-forming disk galaxies, and irregular galaxies (Fig. 1a). We find that galaxies with low star-formation rates contain about 52% of all the stellar mass that is in galaxies more massive than $M_* = 10^9 M_\odot$, which agrees well with observations⁶ (54%–60%). Simulating the formation of realistic disk galaxies, like our own Milky Way, has remained an unsolved problem for more than two decades. The culprit was an angular momentum deficit leading to too high central concentrations, overly massive bulges and unrealistic rotation curves^{7,8}. The fact that our calculation naturally produces a morphological mix of realistic disk galaxies coexisting with a population of ellipticals resolves this long-standing issue. It also shows that previous futile attempts to achieve this were not due to an inherent

¹Department of Physics, Kavli Institute for Astrophysics and Space Research, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA. ²Harvard-Smithsonian Center for Astrophysics, 60 Garden Street, Cambridge, Massachusetts 02138, USA. ³Heidelberg Institute for Theoretical Studies, Schloss-Wolfsbrunnengasse 35, 69118 Heidelberg, Germany. ⁴Zentrum für Astronomie der Universität Heidelberg, ARI, Mönchhofstrasse 12-14, 69120 Heidelberg, Germany. ⁵Kavli Institute for Cosmology, and Institute of Astronomy, Madingley Road, Cambridge CB3 0HA, UK. ⁶Space Telescope Science Institute, 3700 San Martin Drive, Baltimore, Maryland 21218, USA. ⁷Institute for Advanced Study, 1 Einstein Drive, Princeton, New Jersey 08540, USA.

flaw of the Λ CDM paradigm, but rather due to limitations of numerical algorithms and physical modelling.

As our simulation follows the evolution of galaxies starting relatively shortly after the Big Bang, we can construct virtual mock observations that mimic the conditions that the Hubble Space Telescope encounters as it images galaxies across cosmic time in very deep surveys, such as the Ultra Deep Field (UDF), and the continuing Frontier Fields programme. These observations capture a large variety of galaxy luminosities, sizes, colours, morphologies and evolutionary stages, providing remarkable benchmarks for galaxy formation theories. We have constructed a mock UDF and compare it side by side with data from the HST eXtreme Deep Field (XDF) compilation⁹ (Fig. 1b, c). Galaxies in the mock UDF appear strikingly similar to the observed population in terms of number density, colours, sizes and morphologies. Our model is the first hydrodynamic simulation from which a faithful deep UDF-like observation could be constructed, thanks to its combination of large volume and high resolution, along with the new numerical techniques that allow it to reproduce realistic galaxy morphologies.

Satellite galaxies in clusters

This qualitative agreement also extends to many quantitative probes of the distribution and internal structure of galaxies. The abundance and spatial distribution of satellite galaxies forms an important observational test of the Λ CDM paradigm, as it is very sensitive both to details of the hierarchical structure formation process and to galactic-scale

baryonic processes. In fact, two of the most acute challenges to the Λ CDM model are related to satellite galaxies on galactic scales, giving rise to the ‘missing satellite’¹⁰ and ‘too-big-to-fail’ problems¹¹. A particularly taxing problem is the radial distribution of satellite galaxies within galaxy clusters, which are the largest gravitationally bound objects in the Universe.

Recently, large Sloan Digital Sky Survey cluster samples have made it possible to measure accurate radial satellite profiles^{12,13}. Theoretical models have so far struggled to reproduce the shapes and normalizations of these profiles. Semi-analytic models have resorted to *ad hoc* prescriptions for mass-stripping and tuned approximations for satellite orbits because they are missing the gravitational effects of the stellar component. Despite the freedom offered by the adopted coarse parameterizations, most semi-analytic models consistently find radial profiles that are too steep, and in particular overestimate the number of satellites in the inner regions of clusters (radius $r < 100$ – 200 kpc)^{13–15}. Previous hydrodynamic simulations, on the other hand, found inner radial profiles that are too shallow and in most cases could not reproduce the observed normalization^{16–18}, owing both to limited resolution and to an over-production of stars in satellites due to missing physical processes.

For the most massive haloes, which host the largest accumulations of galaxies in the Universe, we have calculated a stacked projected galaxy count profile and compared it to results from a sample¹³ of satellite galaxies with r-band magnitude < -20.5 , for host systems at $0.15 \leq z \leq 0.4$ extracted from 55,121 Sloan Digital Sky Survey groups and clusters ($10^{13.7} M_{\odot} < M_{500, \text{crit}} < 10^{15.0} M_{\odot}$, where $M_{500, \text{crit}}$ is the characteristic mass enclosed in a radius with mean density 500 times the critical density for closure) centred on luminous red galaxies (Fig. 2). Having higher resolution than previous hydrodynamic simulations, and more realistic feedback models that suppress the stellar masses of satellites, we obtain a good agreement with both the observed profile shape and its normalization, as well as with the mean galaxy colour as a function of cluster-centric distance¹⁹. The radial distribution of dark-matter subhaloes from a corresponding dark-matter-only simulation (‘Illustris-Dark’) flattens towards the centre. This difference is due to dissipational processes of galaxy formation that make the stellar component more resistant to tidal disruption close to cluster centres. This directly demonstrates that neglecting baryonic physics causes inaccuracies in the spatial distribution of satellite galaxies which in turn can lead to errors in estimates of galaxy merger rates.

Metals and neutral hydrogen in galaxies

The high mass resolution of the simulation makes it possible to study the internal characteristics of galaxies. It allows us, for example, to make clear predictions for the stellar and gaseous contents of galaxies, and for the baryonic cycle operating between them. Our model predicts the present-day H I mass, $M_{\text{H I}}$, of galaxies in the local Universe, which can be compared to observations as revealed by the Arecibo Legacy FAST ALFA Survey (ALFALFA; Fig. 3a). As stars form out of cold gas which is largely neutral, the H I richness is a good probe of the reservoir of gas available for star formation in a galaxy. The H I-selected samples of ALFALFA are typically biased towards the most gas-rich star-forming galaxies; that is, the major limitation of H I-selected samples is that they will miss the most gas-poor elliptical galaxies.

Our results recover the trend of decreasing H I richness with increasing stellar mass. For the most massive galaxies, the full simulated galaxy sample deviates from the observed mean relation. Many of these massive galaxies have been quenched through AGN feedback and therefore contain little amounts of gas; that is, they would not be detected by ALFALFA given their low H I content. If we focus instead on star-forming galaxies, we find significantly better agreement even at higher masses, and we reproduce the observed H I richness relation over nearly four orders of magnitude in stellar mass. We also include a separate H I richness relation for the satellites of the most massive cluster and compare this to H I measurements of galaxies in the Virgo cluster²⁰ where ram pressure removes gas from infalling galaxies (Fig. 3a). Such hydrodynamic

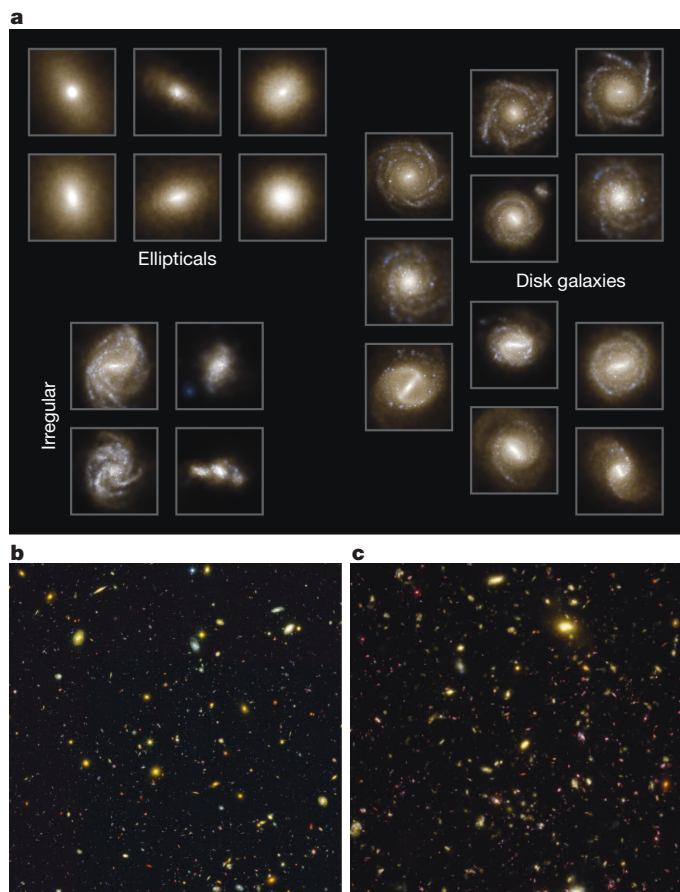


Figure 1 | Mock images of the simulated galaxy population. **a**, Stellar light distributions (g, r, i bands) for a sample of galaxies at $z = 0$ arranged along the classical Hubble sequence for morphological classification. Our simulation produces a variety of galaxy types, ranging from ellipticals to disk galaxies to irregular systems, the last mostly resulting from interactions and mergers. **b**, HST UDF image (2.8 arcmin on a side) in B, Z and H bands convolved with Gaussian point-spread functions of $\sigma = 0.04, 0.08$ and 0.16 arcsec, respectively. **c**, HST mock observation from Illustris.

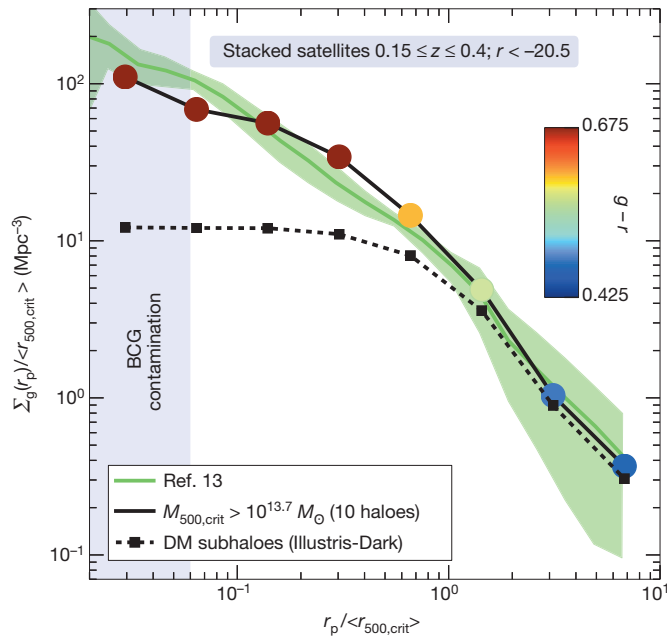


Figure 2 | Projected number density profile of satellite galaxies in galaxy clusters. Symbols show a simulation sample as a function of cluster-centric projected radius, considering 10 haloes more massive than $M_{500,\text{crit}} = 10^{13.7} M_{\odot}$ and all their satellite galaxies brighter than $r < -20.5$ mag within $3 \times R_{500,\text{crit}}$ along the line of sight from the cluster centre. Σ_g is the projected galaxy number density, r_p is the projected distance from the halo centre and $\langle r_{500,\text{crit}} \rangle$ is the mean radius, which encloses $M_{500,\text{crit}}$. The symbol colour reflects the average $(g-r)$ colour for galaxies at that distance (see key). A reddening towards the cluster centre is clearly visible. Observational data¹³ are shown for comparison (green line), while the dashed line gives the subhalo profile of the corresponding dark-matter (DM)-only simulation (normalized to match the satellite profile at large radii). The left blue region marks the area of incompleteness in the observed number density profiles due to obscuration from the brightest cluster galaxy (BCG). The green shaded region marks the observational uncertainties: outside the region of contamination of the BCG the scatter within the observational stack dominates, whereas Poisson errors are shown within that region.

processes are not directly accessible to semi-analytic models but are captured in our simulation. We can therefore recover the observed trend: satellites in cluster environments have a lower H I content compared to the whole population. The predicted difference is not as pronounced as in the observations, but this may well be caused by the fact that our volume does not contain a galaxy cluster as massive as Virgo.

Every dynamical time, a small fraction of the galactic cold neutral gas turns into stars, which inherit the chemical composition of the gas. Stellar metallicities therefore probe different processes intrinsic to galaxy formation: chemical enrichment, feedback, and in- and outflows. Hydrodynamic cosmological simulations have so far been unable to produce correct stellar metallicities; measures such as the mean stellar metallicity or the cosmic density of total metal mass locked up in stars, $\rho_{Z,*}$, were discrepant from the observed values by almost a factor of two^{21,22}. Our model, on the other hand, produces $\rho_{Z,*} = 7.75 \times 10^6 M_{\odot} \text{Mpc}^{-3}$, in agreement with the observed value²³ of $\rho_{Z,*} = (7.1 \pm 2) \times 10^6 M_{\odot} \text{Mpc}^{-3}$.

Recent observations find a relation where more massive galaxies contain a larger proportion of metals^{24–26}. Our predictions agree well with these observations, and, most importantly, the simulation recovers the flattening of the relation above $M_* \approx 10^{11} M_{\odot}$ (Fig. 3b). The observed trend is reproduced over nearly five orders of magnitude in stellar mass. These agreements with observations are driven by the combination of our accurate hydrodynamic scheme together with realistic effective feedback models and a complete stellar evolution prescription, a combination that was lacking in previous studies.

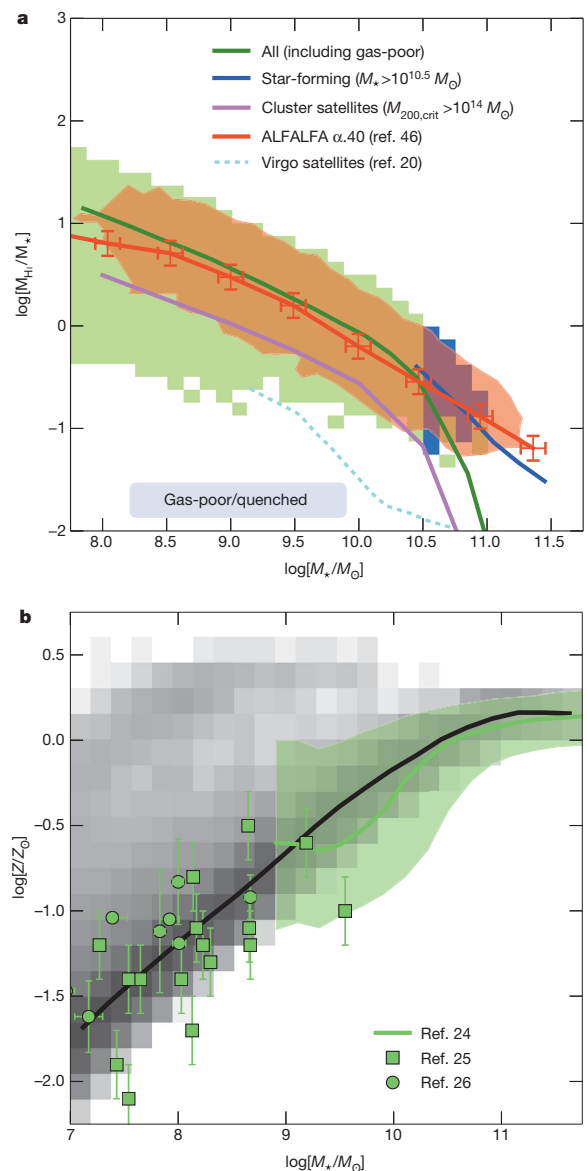


Figure 3 | Neutral hydrogen and metal content of galaxies as a function of their stellar mass. **a**, Predicted H I richness of galaxies (M_{HI} is the mass in neutral hydrogen) compared to the $\alpha.40$ sample of the ALFALFA survey. Filled regions mark bins with more than 20 galaxies (red, ALFALFA; green, all galaxies; blue, star-forming galaxies). Error bars indicate s.e.m. for individual galaxies. A separate H I richness relation for the satellites of the most massive cluster is included and compared to H I measurements of Virgo satellites²⁰ to demonstrate the strong effect of environment on the H I content of satellites. **b**, Stellar metal content of simulated galaxies, Z , in units of solar metallicity, Z_{\odot} (black line and grey shaded squares), compared to observations^{24–26} (green). Reassuringly, the trend of the median relation including the flattening above the stellar mass $M_* \approx 10^{11} M_{\odot}$ is well recovered. The green shaded region represents the s.d. The error bars represent the s.e.m. for individual galaxies.

Large-scale characteristics of neutral hydrogen

The space between galaxies, the intergalactic medium, is filled with low-density, warm gas, mainly composed of ionized hydrogen. Observationally, the intergalactic medium can be probed through the forest of Lyman- α absorption lines in quasar spectra. The corresponding distribution of hydrogen column densities is well-constrained^{27,28} over a wide dynamic range. At high gas densities, hydrogen becomes self-shielded from the ionizing ultraviolet background and forms dense neutral clouds with a distinctive spectral absorption signature, the so-called damped Lyman- α absorbers (DLAs), which are observed mostly at redshift $z \approx 2-5$ and probe initial stages of galaxy formation.

Early hydrodynamic simulations successfully predicted the low-column-density statistics of the Lyman- α forest²⁹, whereas describing the main properties of higher-density absorbers and DLAs correctly only became feasible over the past few years^{30,31}. However, recent numerical calculations have failed to explain the metallicity distribution of DLAs, which is now reasonably well constrained by data³². The outcomes of most simulations gave metallicities for DLAs that are too high or yielded a metallicity distribution that is too broad, in disagreement with observations^{33,34}.

We contrast our predicted H I column density distribution function (CDDF; Fig. 4a) and the metal content of DLAs (Fig. 4b) with observations.

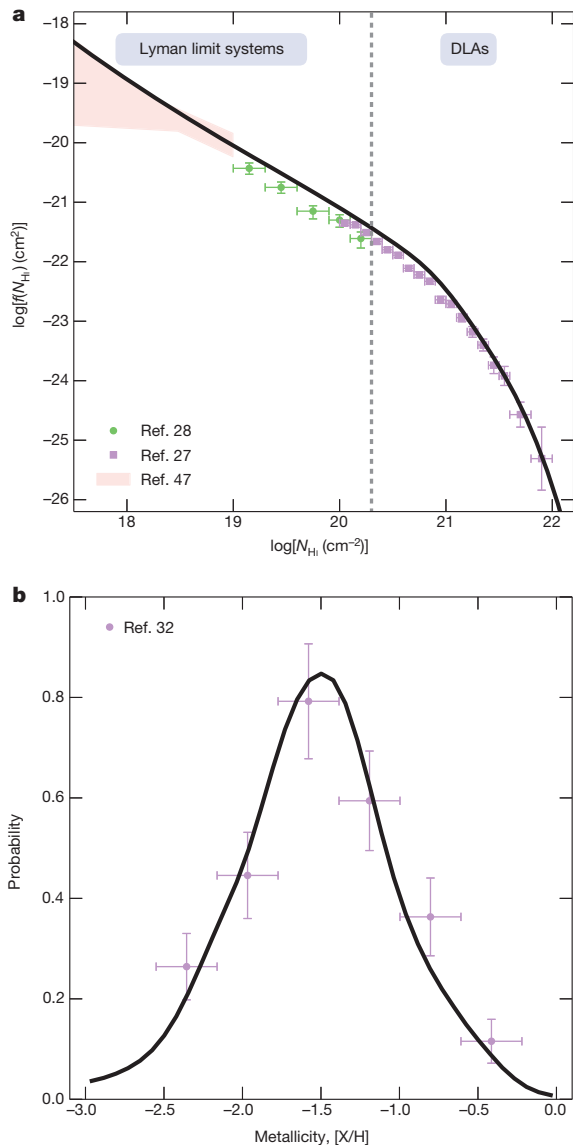


Figure 4 | Large-scale characteristics of neutral hydrogen. **a**, Column density distribution function (CDDF), $f(N_{\text{HI}})$, of neutral hydrogen at $z = 3$ compared to observations^{27,28,47}. The dashed vertical line shows the density threshold separating damped Lyman- α systems (DLAs) and Lyman limit systems. The shaded region shows an estimate for the CDDF constrained by the assumption of a power law fit and the observed incidence of Lyman limit systems. The vertical error bars represent the s.e.m., and the horizontal ones represent the binning. **b**, Probability density function of the DLA metallicity in units of solar metallicity at $z = 3$ compared to observational findings³². The observational vertical error bars show the s.e.m. derived from the number of observed spectra in each metallicity bin. Bin widths have been chosen to be larger than the maximal uncertainty in each individual metallicity measurement (horizontal error bars).

For the DLA metallicities, we compare to an observational compendium³², based on all available quasars between $z = 2$ and $z = 4$, whereas the CDDF data are centred at $z = 3$. The properties of DLAs are sensitive to the balance between gas accretion, outflows and ionization. Probing such a complex interplay of processes is a particular strength of hydrodynamic simulations. The prediction of our model is in remarkable agreement with the observational data, reproducing in detail the shape and location of the metallicity peak. This success is a result of accurate hydrodynamics and modelling of galactic super-winds resulting in metal outflows. Also, we find good agreement between our theoretical predictions and the observed CDDF.

At low redshift, most of the baryons have not yet been detected. Improved instrumental sensitivity may lead to their detection in the foreseeable future³⁵, at a level depending on exactly where they reside, and at what temperature, which is a matter of theoretical debate. On the basis of our model we predict that, at $z = 0$, gas that is not bound to any haloes constitutes 81% of the total baryons in the Universe, but contains only 34% of the heavy metals.

The impact of baryons on dark matter

The discussion thus far has stressed that a direct modelling of baryonic physics is essential to soundly connect cosmological predictions to galaxy observations. Even beyond that, baryonic processes can actually affect and modify the dark-matter distribution, and therefore alter the matter power spectrum $P(k)$. Measuring $P(k)$ as a function of wave-number, k , provides a powerful cosmological probe, because theoretical models can be used to connect the present-day $P(k)$ to the initial power spectrum. Such measurements come, for example, from weak lensing studies, galaxy clustering surveys and the analysis of the Lyman- α forest. However, the interpretation of upcoming weak lensing surveys such as EUCLID, which will measure $P(k)$ on scales of $0.1 h \text{ Mpc}^{-1} < k < 10 h \text{ Mpc}^{-1}$, must consider the impact of baryons in order to achieve

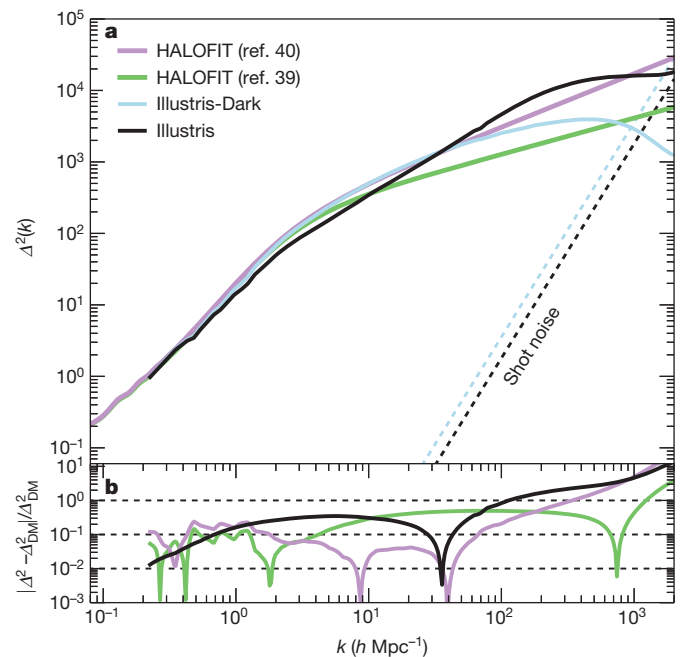


Figure 5 | Nonlinear matter power spectrum. **a**, The dimensionless total matter power spectrum, $\Delta^2(k)$, of the Illustris simulation (black line) differs significantly, owing to baryonic effects, from that of the dark-matter-only counterpart Illustris-Dark (light blue). Analytic fitting models^{39,40} (green and pink) do not provide an adequate description of the hydrodynamic results. The theoretical shot noise level (shown as thin dashed lines) has been subtracted in the measurements. **b**, The relative differences between Illustris and Illustris-Dark highlight the fact that baryonic effects already exceed 1% on scales smaller than $k \approx 1 h \text{ Mpc}^{-1}$. Dashed black lines mark 1%, 10% and 100% relative differences.

the required level of $\sim 1\%$ accuracy^{36,37}. This effect has often been neglected, even though it is not restricted to just the smallest spatial scales.

We have calculated the dimensionless matter power spectrum, $\Delta^2(k) = k^3 P(k)/(2\pi^2)$, along with the corresponding dark-matter-only result (Fig. 5). The dynamic range of our simulation allows us to probe $P(k)$ on a wider range of scales than possible before through a single hydrodynamic simulation. On scales smaller than $k \approx 1h \text{ Mpc}^{-1}$, AGN-driven outflows reduce the total power in a scale-dependent way by up to $\sim 30\%$ – 40% compared to the dark-matter-only prediction. This impact is even larger than found in previous studies³⁸, which is most probably related to the strong radio-mode AGN feedback in our simulation needed to match the stellar mass content of massive haloes. On smaller scales, gas cooling processes become important and enhance the power on scales smaller than $k \approx 100h \text{ Mpc}^{-1}$. Here, the power spectrum can deviate by a factor of a few compared to collisionless results. Measuring this fundamental statistic precisely hence requires high-resolution hydrodynamic simulations in large volumes; it cannot be done by dark-matter-only simulations or semi-analytic modelling. We also present in Fig. 5 the predictions of commonly employed empirical fitting models for the nonlinear evolution^{39,40} of $P(k)$. Those models have been calibrated on dark-matter-only simulations and clearly fail to describe the results of full hydrodynamic calculations, rendering their use impractical for the precision requirements of upcoming surveys.

Looking ahead

Although our simulation provides a significant step forward in modelling galaxy formation by reproducing simultaneously many disparate observations on large and small scales, there are still outstanding problems. One such problem lies in the formation of low-mass galaxies: our simulation tends to build up the stellar mass of low-mass galaxies below $M_* \approx 10^{10} M_\odot$ too early, resulting in stellar populations that are too old, with mean ages a factor of two to three larger than observed. This tension is shared by semi-analytic models and other recent hydrodynamic simulations alike, pointing towards an open problem in low-mass galaxy formation⁴¹. It remains to be seen whether new stellar feedback models that, for example, include effects of radiation pressure on dust can resolve this issue⁴². It will clearly be challenging to test new schemes that also directly treat the stellar radiation fields with statistically meaningful samples of galaxies, as this poses extremely high computational demands that go significantly beyond what was achieved in the present work. Nevertheless, such new generations of large-scale high-resolution hydrodynamic simulations might become feasible within the next decade.

METHODS SUMMARY

The equations of gravity and hydrodynamics are evolved using the moving-mesh code AREPO⁵ combined with a galaxy formation model which includes⁴³: gas cooling with radiative self-shielding corrections; star formation; energetic feedback from growing SMBHs and exploding supernovae; stellar evolution with associated chemical enrichment and stellar mass loss; and radiation proximity effects for AGNs. Our AGN feedback consists of three components: thermal quasar-mode feedback; thermal-mechanical radio-mode feedback; and radiative feedback. The efficiency of physical processes below the resolution scale are treated in parameterized form and have been calibrated to reproduce the observed global efficiency of star formation. Our simulation assumes a Λ CDM cosmology with the parameters $\Omega_m = 0.2726$, $\Omega_\Lambda = 0.7274$, $\Omega_b = 0.0456$, $\sigma_8 = 0.809$, $n_s = 0.963$ and $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$, with $h = 0.704$. Here Ω_m , Ω_Λ , and Ω_b are respectively the mass/energy density contributions from matter, the cosmological constant and baryons, h encodes the Hubble expansion rate at redshift zero, n_s is the spectral index of the primordial power spectrum and σ_8 is the root mean squared amplitude of mass fluctuations in $8h^{-1} \text{ Mpc}$ spheres linearly extrapolated to redshift zero. Initial conditions are generated at $z = 127$ in a periodic box with a side length of $75h^{-1} \text{ Mpc} \approx 106.5 \text{ Mpc}$. The initial gas temperature at $z = 127$ is set to 245 K . We have also performed a second simulation, Illustris-Dark, which does not include baryons and the related feedback, but is otherwise identical to Illustris. Dark-matter haloes were identified in an on-the-fly manner during the simulation for each snapshot using a friends-of-friends (FOF) algorithm with a linking length of 0.2 times the mean particle separation. The minimum dark-matter particle number was set to 32 for the FOF

identification. Non-dark-matter particles are attached to these FOF primaries in a secondary linking stage⁴⁴. Subsequently, gravitationally bound substructures are identified using the SUBFIND algorithm^{44,45}. 16 million CPU hours were needed to evolve the simulation from the starting redshift $z = 127$ to $z = 0$, using 8,192 cores and an equal number of MPI-ranks. An additional 3 million CPU hours were spent on carrying out the on-the-fly galaxy identification with the SUBFIND algorithm.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 January; accepted 7 April 2014.

- Planck Collaboration. Planck 2013 results. XVI. Cosmological parameters. Preprint at <http://arxiv.org/abs/1303.5076> (2013).
- Katz, N. Dissipational galaxy formation. II — Effects of star formation. *Astrophys. J.* **391**, 502–517 (1992).
- Schaye, J. *et al.* The physics driving the cosmic star formation history. *Mon. Not. R. Astron. Soc.* **402**, 1536–1560 (2010).
- Brook, C. B., Stinson, G., Gibson, B. K., Wadsley, J. & Quinn, T. MaGICC discs: matching observed galaxy relationships over a wide stellar mass range. *Mon. Not. R. Astron. Soc.* **424**, 1275–1283 (2012).
- Springel, V. *et al.* Galilean-invariant cosmological hydrodynamical simulations on a moving mesh. *Mon. Not. R. Astron. Soc.* **401**, 791–851 (2010).
- Moustakas, J. *et al.* PRIMUS: Constraints on star formation quenching and galaxy merging, and the evolution of the stellar mass function from $z = 0$ –1. *Astrophys. J.* **767**, 50 (2013).
- Navarro, J. F. & Steinmetz, M. Dark halo and disk galaxy scaling laws in hierarchical universes. *Astrophys. J.* **538**, 477–488 (2000).
- Scannapieco, C. *et al.* The Aquila comparison project: the effects of feedback and numerical methods on simulations of galaxy formation. *Mon. Not. R. Astron. Soc.* **423**, 1726–1749 (2012).
- Illingworth, G. D. *et al.* The HST eXtreme Deep Field (XDF): combining all ACS and WFC3/IR data on the HUDF region into the deepest field ever. *Astrophys. J. Suppl. Ser.* **209**, 6 (2013).
- Moore, B. *et al.* Dark matter substructure within galactic halos. *Astrophys. J.* **524**, L19–L22 (1999).
- Boylan-Kolchin, M., Bullock, J. S. & Kaplinghat, M. Too big to fail? The puzzling darkness of massive Milky Way subhaloes. *Mon. Not. R. Astron. Soc.* **415**, L40–L44 (2011).
- Hansen, S. M. *et al.* Measurement of galaxy cluster sizes, radial profiles, and luminosity functions from SDSS photometric data. *Astrophys. J.* **633**, 122–137 (2005).
- Budzynski, J. M., Koposov, S. E., McCarthy, I. G., McGee, S. L. & Belokurov, V. The radial distribution of galaxies in groups and clusters. *Mon. Not. R. Astron. Soc.* **423**, 104–121 (2012).
- Quilis, V. & Trujillo, I. Satellites around massive galaxies since $z \sim 2$: confronting the millennium simulation with observations. *Astrophys. J.* **752**, L19 (2012).
- Tal, T. *et al.* Galaxy environments over cosmic time: the non-evolving radial galaxy distributions around massive galaxies since $z = 1.6$. *Astrophys. J.* **769**, 31 (2013).
- Nagai, D. & Kravtsov, A. V. The radial distribution of galaxies in Λ cold dark matter clusters. *Astrophys. J.* **618**, 557–568 (2005).
- Saro, A. *et al.* Properties of the galaxy population in hydrodynamical simulations of clusters. *Mon. Not. R. Astron. Soc.* **373**, 397–410 (2006).
- Libeskind, N. I., Cole, S., Frenk, C. S., Okamoto, T. & Jenkins, A. Satellite systems around galaxies in hydrodynamic simulations. *Mon. Not. R. Astron. Soc.* **374**, 16–28 (2007).
- Hansen, S. M., Sheldon, E. S., Wechsler, R. H. & Koester, B. P. The galaxy content of SDSS clusters and groups. *Astrophys. J.* **699**, 1333–1353 (2009).
- Cortese, L., Catinella, B., Boissier, S., Boselli, A. & Heinis, S. The effect of the environment on the H I scaling relations. *Mon. Not. R. Astron. Soc.* **415**, 1797–1806 (2011).
- Davé, R. & Oppenheimer, B. D. The enrichment history of baryons in the Universe. *Mon. Not. R. Astron. Soc.* **374**, 427–435 (2007).
- Wiersma, R. P. C., Schaye, J. & Theuns, T. The effect of variations in the input physics on the cosmic distribution of metals predicted by simulations. *Mon. Not. R. Astron. Soc.* **415**, 353–371 (2011).
- Gallazzi, A., Brinchmann, J., Charlot, S. & White, S. D. M. A census of metals and baryons in stars in the local Universe. *Mon. Not. R. Astron. Soc.* **383**, 1439–1458 (2008).
- Gallazzi, A., Charlot, S., Brinchmann, J., White, S. D. M. & Tremonti, C. A. The ages and metallicities of galaxies in the local universe. *Mon. Not. R. Astron. Soc.* **362**, 41–58 (2005).
- Woo, J., Courteau, S. & Dekel, A. Scaling relations and the fundamental line of the local group dwarf galaxies. *Mon. Not. R. Astron. Soc.* **390**, 1453–1469 (2008).
- Kirby, E. N. *et al.* The universal stellar mass–stellar metallicity relation for dwarf galaxies. *Astrophys. J.* **779**, 102 (2013).
- Noterdaeme, P., Petitjean, P., Ledoux, C. & Srianand, R. Evolution of the cosmological mass density of neutral gas from Sloan Digital Sky Survey II – Data Release 7. *Astron. Astrophys.* **505**, 1087–1098 (2009).
- Zafar, T. *et al.* The ESO UVES advanced data products quasar sample. II. Cosmological evolution of the neutral gas mass density. *Astron. Astrophys.* **556**, A141 (2013).
- Hernquist, L., Katz, N., Weinberg, D. H. & Miralda-Escudé, J. The Lyman-alpha forest in the cold dark matter model. *Astrophys. J.* **457**, L51 (1996).

30. Pontzen, A. *et al.* Damped Lyman α systems in galaxy formation simulations. *Mon. Not. R. Astron. Soc.* **390**, 1349–1371 (2008).
31. Altay, G., Theuns, T., Schaye, J., Booth, C. M. & Dalla Vecchia, C. The impact of different physical processes on the statistics of Lyman-limit and damped Lyman α absorbers. *Mon. Not. R. Astron. Soc.* **436**, 2689–2707 (2013).
32. Rafelski, M., Wolfe, A. M., Prochaska, J. X., Neeleman, M. & Mendez, A. J. Metallicity evolution of damped Ly α systems out to $z \sim 5$. *Astrophys. J.* **755**, 89 (2012).
33. Fumagalli, M. *et al.* Absorption-line systems in simulated galaxies fed by cold streams. *Mon. Not. R. Astron. Soc.* **418**, 1796–1821 (2011).
34. Cen, R. The nature of damped Ly α systems and their hosts in the standard cold dark matter universe. *Astrophys. J.* **748**, 121 (2012).
35. Bregman, J. N. The search for the missing baryons at low redshift. *Annu. Rev. Astron. Astrophys.* **45**, 221–259 (2007).
36. Huterer, D. & Takada, M. Calibrating the nonlinear matter power spectrum: requirements for future weak lensing surveys. *Astropart. Phys.* **23**, 369–376 (2005).
37. Laureijs, R. *et al.* Euclid assessment study report for the ESA cosmic visions. Preprint at <http://arxiv.org/abs/0912.0914> (2009).
38. van Daalen, M. P., Schaye, J., Booth, C. M. & Dalla Vecchia, C. The effects of galaxy formation on the matter power spectrum: a challenge for precision cosmology. *Mon. Not. R. Astron. Soc.* **415**, 3649–3665 (2011).
39. Smith, R. E. *et al.* Stable clustering, the halo model and non-linear cosmological power spectra. *Mon. Not. R. Astron. Soc.* **341**, 1311–1332 (2003).
40. Takahashi, R., Sato, M., Nishimichi, T., Taruya, A. & Oguri, M. Revising the Halofit model for the nonlinear matter power spectrum. *Astrophys. J.* **761**, 152 (2012).
41. Weinmann, S. M. *et al.* A fundamental problem in our understanding of low-mass galaxy evolution. *Mon. Not. R. Astron. Soc.* **426**, 2797–2812 (2012).
42. Hopkins, P. F. *et al.* Galaxies on FIRE (Feedback In Realistic Environments): stellar feedback explains cosmologically inefficient star formation. Preprint at <http://arxiv.org/abs/1311.2073> (2013).
43. Vogelsberger, M. *et al.* A model for cosmological simulations of galaxy formation physics. *Mon. Not. R. Astron. Soc.* **436**, 3031–3067 (2013).
44. Dolag, K., Borgani, S., Murante, G. & Springel, V. Substructures in hydrodynamical cluster simulations. *Mon. Not. R. Astron. Soc.* **399**, 497–514 (2009).
45. Springel, V., White, S. D. M., Tormen, G. & Kauffmann, G. Populating a cluster of galaxies — I. Results at $z=0$. *Mon. Not. R. Astron. Soc.* **328**, 726–750 (2001).
46. Huang, S., Haynes, M. P., Giovanelli, R. & Brinchmann, J. The Arecibo Legacy Fast ALFA Survey: the galaxy population detected by ALFALFA. *Astrophys. J.* **756**, 113 (2012).
47. Prochaska, J. X., O'Meara, J. M. & Worseck, G. A definitive survey for Lyman limit systems at $z \sim 3.5$ with the Sloan Digital Sky Survey. *Astrophys. J.* **718**, 392–416 (2010).

Acknowledgements V.S. acknowledges support from the DFG Research Centre SFB-881 'The Milky Way System' through project A1, and from the European Research Council under ERC-StG EXAGAL-308037. G.S. acknowledges support from the HST grants programme, no. HST-AR-12856.01-A. Support for program no. 12856 was provided by NASA through a grant from the Space Telescope Science Institute, which is operated by the Association of Universities for Research in Astronomy, Inc., under NASA contract NAS 5-26555. L.H. acknowledges support from NASA grant NNX12AC67G and NSF grant AST-1312095. D.X. acknowledges support from the Alexander von Humboldt Foundation. S.B. was supported by NSF grant AST-0907969. The Illustris simulation was run on the CURIE supercomputer at CEA/France as part of PRACE project RA0844, and the SuperMUC computer at the Leibniz Computing Centre, Germany, as part of GCS-project pr85je. Further simulations were run on the Harvard Odyssey and Cfa/ITC clusters, the Ranger and Stampede supercomputers at the Texas Advanced Computing Center through XSEDE, and the Kraken supercomputer at Oak Ridge National Laboratory through XSEDE. Figure 1b is based on observations made with the NASA/ESA Hubble Space Telescope. These data were obtained from the Mikulski Archive for Space Telescopes (MAST) at the Space Telescope Science Institute (STScI). These observations were associated with programs 9,352, 9,425, 9,488, 9,575, 9,793, 9,978, 10,086, 10,189, 10,258, 10,340, 10,530, 11,359, 11,563, 12,060, 12,061, 12,062, 12,099 and 12,177, and compiled for the Hubble eXtreme Deep Field data release version 1.0 (<http://archive.stsci.edu/prepds/xdfl/>). Support for MAST for non-HST data is provided by the NASA Office of Space Science via grant NNX13AC07G and by other grants and contracts.

Author Contributions M.V., L.H., D.S., V.S. and S.G. conceived and planned the project. M.V., S.G., D.S. and P.T. developed the galaxy formation model. V.S. developed the AREPO code. M.V. generated initial conditions. V.S., M.V. and S.G. ran the simulations. M.V. performed the main analysis. G.S. and P.T. constructed the mock images. S.B. provided statistics of the inter-galactic medium. D.X. and D.N. provided post-processing tools. M.V., S.G., V.S., P.T. and L.H. interpreted the results. M.V. and S.G. wrote the manuscript with contributions from co-authors.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to M.V. (mvogelsb@mit.edu).

METHODS

Simulation code. Our simulation code, AREPO, uses an unstructured Voronoi tessellation of the simulation volume, where the mesh-generating points of this tessellation are moved with the gas flow. The adaptive mesh is used to solve the equations of ideal hydrodynamics with a finite volume approach using a second-order unsplit Godunov scheme with an exact Riemann solver. This approach is under most circumstances superior to traditional smoothed particle hydrodynamics (SPH), and also to Eulerian adaptive mesh refinement (AMR)^{48–51}. This scheme naturally produces extended disk galaxies without invoking extreme forms of stellar feedback or star formation⁵², which was a major problem of previous galaxy formation simulations. The gravity calculation employs a Tree-PM scheme⁵³, where long-range forces are determined with a particle-mesh method (PM) while short-range forces are computed via a hierarchical tree algorithm⁵⁴.

Galaxy formation physics. Our simulation accounts for a variety of astrophysical processes known to be relevant for galaxy formation⁴³.

Gas cooling rates are calculated as a function of gas density, temperature, metallicity, the radiation fields of active galactic nuclei and the spatially uniform but time-dependent ionizing background radiation from galaxies and quasars⁵⁵, which completes H I reionization at a redshift of $z \approx 6$. We use a self-consistent calculation of the primordial cooling and complement it with the cooling contribution of the metals, based on pre-calculated cooling rate tables using CLOUDY⁵⁶. All cooling rates include self-shielding corrections⁵⁷.

We employ a sub-resolution model of the interstellar medium to achieve numerical closure below our resolution scale. Gas with hydrogen number density above 0.13 cm^{-3} follows an effective equation of state with a stochastic prescription for star formation following the Kennicutt-Schmidt law⁵⁸ and adopting a Chabrier initial mass function⁵⁹. The effective equation of state assumes that the interstellar medium has a two-phase structure that is predominantly composed of cold clouds embedded in a tenuous, supernova-heated phase⁶⁰.

Once stellar populations are born, they can lose mass, for example through stellar winds or supernovae. This mass is returned to the gas phase and enriches the gas surrounding stellar populations. We track the evolution of stars and model supernovae of type Ia, type II and the asymptotic giant branch phases of stars. We trace the evolution of nine elements in total (H, He, C, N, O, Ne, Mg, Si, Fe), each advected as a passive scalar.

Stellar feedback is realized through a kinetic wind scheme with a velocity scaling based on the local one-dimensional dark-matter velocity dispersion ($3.7\sigma_{\text{DM,1D}}$), and a mass loading inferred from energy conservation assuming $1.09 \times 10^{51} \text{ erg}$ per type II supernova. We use a sub-grid metal-loading scheme that regulates the degree of wind enrichment such that 40% of the local interstellar medium metals are ejected by supernova-driven galactic winds. This is required to simultaneously reproduce the stellar mass content of low-mass haloes and their gas oxygen abundances⁶¹.

We include procedures for supermassive black hole (SMBH) seeding, accretion and merging⁶². Feedback from SMBHs operates in either a quasar-mode or a radio-mode, depending on their accretion rate⁶³. In addition, a prescription for radiative SMBH feedback is included that modifies the ionization state and hence the net cooling rate of nearby gas. SMBHs are seeded in friends-of-friends groups more massive than $7.1 \times 10^{10} M_{\odot}$ with a seed mass of $1.4 \times 10^5 M_{\odot}$.

The free parameters of our model are set to physically plausible values and have been adjusted within the allowed range to roughly reproduce the relation between mean stellar mass and halo mass inferred from abundance matching analysis. The resulting parameter settings have been tested on smaller-scale simulations⁴³ and high-resolution zoom-in simulations of individual Milky Way-like haloes⁶⁴.

Initial conditions. To create initial conditions we use the Boltzmann code CAMB^{65,66} to compute the linear power spectrum of a Λ CDM cosmology with the parameters $\Omega_m = 0.2726$, $\Omega_{\Lambda} = 0.7274$, $\Omega_b = 0.0456$, $\sigma_8 = 0.809$, $n_s = 0.963$ and $H_0 = 100h \text{ km s}^{-1} \text{ Mpc}^{-1}$ with $h = 0.704$. These parameters are consistent with the latest Wilkinson Microwave Anisotropy Probe (WMAP)-9 measurements⁶⁷, but slightly offset from the first year results of the Planck mission¹. However, a recent re-analysis of the Planck data found parameters more consistent with pre-Planck cosmic microwave background analyses and astronomical observations⁶⁸.

We create a random realization of this cosmology in periodic boxes with a side length of $75h^{-1} \text{ Mpc} \approx 106.5 \text{ Mpc}$, starting from an initial ‘glass-like’ particle configuration⁶⁹ composed of one thousand 182^3 particle tiles. We employ a $3,640^3$ fast Fourier transform to calculate the displacement field and use Lagrangian perturbation theory (Zel’dovich approximation⁷⁰) to displace particles, and we de-convolve the input power spectrum for smoothing effects due to the interpolation off this grid. Initial conditions are generated at $z = 127$ with mesh-generating points added to the initial conditions by splitting each original particle into a dark matter and gas cell pair, displacing them with respect to one another such that two interleaved grids are formed, keeping the centre-of-mass of each pair fixed. The initial gas temperature at $z = 127$ is set to 245 K based on a RECFAST^{71,72} calculation. We have generated 100 different random fields and inspected their power spectra and mass

functions at $z = 0$ to make sure that we do not simulate an unusual or extreme density field that is dominated by a few large clusters or voids due to cosmic variance.

Simulation details. The simulation volume contains initially 6,028,568,000 hydrodynamic cells and the same number of dark-matter particles resulting in a dark-matter mass resolution of $6.26 \times 10^6 M_{\odot}$, and a baryonic mass resolution of $1.26 \times 10^6 M_{\odot}$. The gravitational softening length of dark-matter particles is fixed in comoving coordinates ($\epsilon_{\text{DM}} = 1h^{-1} \text{ kpc}$). For baryonic particles (stars and SMBHs), we limit the softening length to a maximum physical scale ($\epsilon_{\text{baryon}} = 0.5h^{-1} \text{ kpc}$). Gas cells use an adaptive softening length tied to their cell radius with a floor given by the softening length of the collisionless baryonic particles. We employ a (de-) refinement scheme which keeps the cell masses typically within a factor of two of a specified target mass set to $1.26 \times 10^6 M_{\odot}$, and a regularization scheme steering the mesh towards a computationally efficient centroidal configuration^{54,48}. The smallest cells in Illustris have a typical extent of 48 pc. For the least massive cells we achieve a mass resolution of $1.5 \times 10^4 M_{\odot}$.

Galaxy identification. Dark-matter haloes were identified in an on-the-fly manner during the simulation for each snapshot using a friends-of-friends (FOF) algorithm⁷³ with a linking length of 0.2 times the mean particle separation. The minimum dark-matter particle number was set to 32 for the FOF identification. Non-dark-matter particles are attached to these FOF primaries in a secondary linking stage⁴⁴. Subsequently, gravitationally bound substructures are identified using the SUBFIND algorithm^{44,45}. We derive the various galaxy properties from the gravitationally bound mass that is contained within a radius r_* that equals twice the stellar half-mass radius of each SUBFIND (sub)halo. Using this definition, the galactic stellar mass does not significantly differ from the total stellar mass for low-mass systems, but some of the intra-cluster light for massive systems is excluded. We have checked this definition against surface brightness cuts in different bands and find it to give similar results to more elaborate methods for excluding intra-cluster light. We have extended this radius by 50% for measuring the galactic H I masses. Stellar population synthesis models⁷⁴ were used to associate the stars in our simulation with observable broad band luminosities. Producing stellar images of galaxies requires assigning colour values to specific bands. Specifically, we make an RGB mapping of the (g,r,i) bands using a commonly employed asinh scaling⁷⁵.

48. Vogelsberger, M., Sijacki, D., Kereš, D., Springel, V. & Hernquist, L. Moving mesh cosmology: numerical techniques and global statistics. *Mon. Not. R. Astron. Soc.* **425**, 3024–3057 (2012).
49. Sijacki, D., Vogelsberger, M., Kereš, D., Springel, V. & Hernquist, L. Moving mesh cosmology: the hydrodynamics of galaxy formation. *Mon. Not. R. Astron. Soc.* **424**, 2999–3027 (2012).
50. Kereš, D., Vogelsberger, M., Sijacki, D., Springel, V. & Hernquist, L. Moving-mesh cosmology: characteristics of galaxies and haloes. *Mon. Not. R. Astron. Soc.* **425**, 2027–2048 (2012).
51. Genel, S. *et al.* Following the flow: tracer particles in astrophysical fluid simulations. *Mon. Not. R. Astron. Soc.* **435**, 1426–1442 (2013).
52. Torrey, P., Vogelsberger, M., Sijacki, D., Springel, V. & Hernquist, L. Moving-mesh cosmology: properties of gas discs. *Mon. Not. R. Astron. Soc.* **427**, 2224–2238 (2012).
53. Xu, G. A new parallel N-body gravity solver: TPM. *Astrophys. J. Suppl. Ser.* **98**, 355 (1995).
54. Barnes, J. & Hut, P. A hierarchical $O(N \log N)$ force-calculation algorithm. *Nature* **324**, 446–449 (1986).
55. Faucher-Giguère, C.-A., Lidz, A., Zaldarriaga, M. & Hernquist, L. A new calculation of the ionizing background spectrum and the effects of He II reionization. *Astrophys. J.* **703**, 1416–1443 (2009).
56. Ferland, G. J. *et al.* CLOUDY 90: numerical simulation of plasmas and their spectra. *Publ. Astron. Soc. Pacif.* **110**, 761–778 (1998).
57. Rahmati, A., Pawlik, A. H., Raicevic, M. & Schaye, J. On the evolution of the H I column density distribution in cosmological simulations. *Mon. Not. R. Astron. Soc.* **430**, 2427–2445 (2013).
58. Kennicutt, R. C., Jr. The star formation law in galactic disks. *Astrophys. J.* **344**, 685–703 (1989).
59. Chabrier, G. Galactic stellar and substellar initial mass function. *Publ. Astron. Soc. Pacif.* **115**, 763–795 (2003).
60. Springel, V. & Hernquist, L. Cosmological smoothed particle hydrodynamics simulations: a hybrid multiphase model for star formation. *Mon. Not. R. Astron. Soc.* **339**, 289–311 (2003).
61. Zahid, H. J. *et al.* Empirical constraints for the magnitude and composition of galactic winds. *Astrophys. Space Sci.* **349**, 873–879 (2014).
62. Di Matteo, T., Springel, V. & Hernquist, L. Energy input from quasars regulates the growth and activity of black holes and their host galaxies. *Nature* **433**, 604–607 (2005).
63. Sijacki, D., Springel, V., Di Matteo, T. & Hernquist, L. A unified model for AGN feedback in cosmological simulations of structure formation. *Mon. Not. R. Astron. Soc.* **380**, 877–900 (2007).
64. Marinacci, F., Pakmor, R. & Springel, V. The formation of disc galaxies in high-resolution moving-mesh cosmological simulations. *Mon. Not. R. Astron. Soc.* **437**, 1750–1775 (2014).
65. Seljak, U. & Zaldarriaga, M. A line-of-sight integration approach to cosmic microwave background anisotropies. *Astrophys. J.* **469**, 437 (1996).

66. Lewis, A. & Challinor, A. CAMB: Code for Anisotropies in the Microwave Background. *Astrophysics Source Code Library* <http://asterisk.apod.com/wp/> (2011).
67. Hinshaw, G. *et al.* Nine-year Wilkinson Microwave Anisotropy Probe (WMAP) observations: cosmological parameter results. *Astrophys. J. Suppl. Ser.* **208**, 19 (2013).
68. Spergel, D., Flauger, R. & Hlozek, R. Planck data reconsidered. Preprint at <http://arxiv.org/abs/1312.3313> (2013).
69. White, S. D. M. in *Cosmology and Large Scale Structure* (eds Schaeffer, R., Silk, J., Spiro, M. & Zinn-Justin, J.) 349 (Elsevier, 1996).
70. Zel'dovich, Y. B. Gravitational instability: an approximate theory for large density perturbations. *Astron. Astrophys.* **5**, 84–89 (1970).
71. Seager, S., Sasselov, D. D. & Scott, D. A new calculation of the recombination epoch. *Astrophys. J.* **523**, L1–L5 (1999).
72. Seager, S., Sasselov, D. D. & Scott, D. RECFast: Calculate the Recombination History of the Universe. *Astrophysics Source Code Library* <http://asterisk.apod.com/wp/> (2011).
73. Davis, M., Efstathiou, G., Frenk, C. S. & White, S. D. M. The evolution of large-scale structure in a universe dominated by cold dark matter. *Astrophys. J.* **292**, 371–394 (1985).
74. Bruzual, G. & Charlot, S. Stellar population synthesis at the resolution of 2003. *Mon. Not. R. Astron. Soc.* **344**, 1000–1028 (2003).
75. Lupton, R. *et al.* Preparing red-green-blue images from CCD data. *Publ. Astron. Soc. Pacif.* **116**, 133–137 (2004).

FXR is a molecular target for the effects of vertical sleeve gastrectomy

Karen K. Ryan¹, Valentina Tremaroli², Christoffer Clemmensen^{1,3}, Petia Kovatcheva-Datchary², Andriy Myronovych⁴, Rebekah Karns⁵, Hilary E. Wilson-Pérez¹, Darleen A. Sandoval¹, Rohit Kohli⁴, Fredrik Bäckhed^{2,6} & Randy J. Seeley¹

Bariatric surgical procedures, such as vertical sleeve gastrectomy (VSG), are at present the most effective therapy for the treatment of obesity, and are associated with considerable improvements in co-morbidities, including type-2 diabetes mellitus. The underlying molecular mechanisms contributing to these benefits remain largely undetermined, despite offering the potential to reveal new targets for therapeutic intervention. Substantial changes in circulating total bile acids are known to occur after VSG. Moreover, bile acids are known to regulate metabolism by binding to the nuclear receptor FXR (farnesoid-X receptor, also known as NR1H4). We therefore examined the results of VSG surgery applied to mice with diet-induced obesity and targeted genetic disruption of FXR. Here we demonstrate that the therapeutic value of VSG does not result from mechanical restriction imposed by a smaller stomach. Rather, VSG is associated with increased circulating bile acids, and associated changes to gut microbial communities. Moreover, in the absence of FXR, the ability of VSG to reduce body weight and improve glucose tolerance is substantially reduced. These results point to bile acids and FXR signalling as an important molecular underpinning for the beneficial effects of this weight-loss surgery.

Perhaps surprisingly, the most effective and durable therapies for the treatment of obesity involve surgical, rather than pharmacological or behavioural, intervention. In contrast to the modest impact of diet and exercise, most individuals that lose weight with bariatric surgery maintain reduced levels of body fat for many years¹. This is associated with considerable improvements in co-morbidities. Notably, 40% of obese type-2 diabetes mellitus patients that undergo bariatric surgery achieve full remission within one year². Despite its efficacy, however, surgery is not an attractive therapeutic option for many individuals suffering from obesity and its consequences. Therefore it is both scientifically and clinically imperative that we identify molecular mechanisms responsible for weight loss and other metabolic improvements so we can target affected pathways in a less-invasive manner.

In the bariatric procedure VSG, approximately 80% of the stomach is removed along the greater curvature. This creates a gastric 'sleeve' in continuity with the oesophagus and duodenum. VSG induces loss of body weight and fat mass, and improves glucose tolerance in humans and in rodents^{2,3}. Importantly, the efficacy of VSG is comparable to the more complex Roux-en-Y gastric bypass (RYGB), which not only creates a smaller stomach but also involves surgical re-routing of the small intestine^{4,5}. A recent randomized controlled trial found that VSG and RYGB produced nearly the same remission of type-2 diabetes mellitus after one year, and that both procedures were significantly superior to medical management².

Conventional wisdom holds that bariatric surgeries, including VSG and RYGB, lead to weight loss and resolution of type-2 diabetes mellitus directly as a result of reducing the ability of the stomach to physically accommodate a meal. That is, the stomach is made smaller, perhaps imposing a mechanical restriction on the amount of food that can be consumed at one time. Moreover, in RYGB, the flow of nutrients and biliopancreatic secretions is re-routed, perhaps limiting macronutrient absorption. However, growing evidence indicates that restriction and

malabsorption are not the primary mechanisms driving metabolic improvements after bariatric surgery⁶. As just one example, a substantial proportion of diabetic patients are able to stop taking their medications within days of surgery, before substantial weight loss has occurred². Given the short time frame, some controversy exists about the mechanisms underlying this dramatic outcome. Some hypothesize that acute improvements are independent of weight loss and result from changes in gut physiology, including altered neurohumoral signalling or altered microbial ecology. In contrast, others suggest these benefits are the result of the hypocaloric post-operative milieu and rapid weight loss^{7,8}.

Among the changes to gut physiology that occur following bariatric surgery is altered enterohepatic circulation of bile acids. Both RYGB and VSG are associated with a significant increase in circulating total bile acids in humans and in rodent models^{9,10}. It is now clear that, in addition to aiding the mechanical digestion and absorption of lipids, bile acids also bind to FXR to function as signalling molecules contributing to the regulation of various metabolic processes. In light of this role, we hypothesized that FXR-signalling links altered bile acid homeostasis to important post-operative changes in metabolism and gut microbial communities, thereby contributing to the maintenance of weight loss and improvements in glucose control observed following VSG. To test this hypothesis, we applied VSG surgery to diet-induced obese mice with targeted genetic disruption of FXR and their wild-type littermates.

Unbiased pathway analysis

We used sequencing of messenger RNA (mRNA-seq), together with subsequent unbiased pathway analysis, to identify key biological pathways that were altered in the distal small intestine following VSG (Fig. 1). Consistent with our recent report that VSG significantly alters the expression of hepatic genes involved in lipid and bile acid metabolism⁹, the pathway 'Nuclear receptors in lipid metabolism and toxicity' emerged as one of the top pathways enriched in genes differentially regulated (fold

¹Department of Internal Medicine, Division of Endocrinology, Diabetes and Metabolism, University of Cincinnati, Cincinnati, Ohio 45237, USA. ²Wallenberg Laboratory, Department of Molecular and Clinical Medicine and Sahlgrenska Center for Cardiovascular and Metabolic Research, University of Gothenburg, S-413 45 Gothenburg, Sweden. ³Department of Drug Design and Pharmacology, Faculty of Health and Medical Sciences, University of Copenhagen, Universitetsparken 2, DK-2100 Copenhagen, Denmark. ⁴Department of Pediatrics, Division of Gastroenterology, Hepatology, and Nutrition, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ⁵Division of Biomedical Informatics, Cincinnati Children's Hospital Medical Center, Cincinnati, Ohio 45229, USA. ⁶Novo Nordisk Foundation Center for Basic Metabolic Research, Section for Metabolic Receptology and Enterendocrinology, Faculty of Health Sciences, University of Copenhagen, Copenhagen, DK-2200, Denmark.

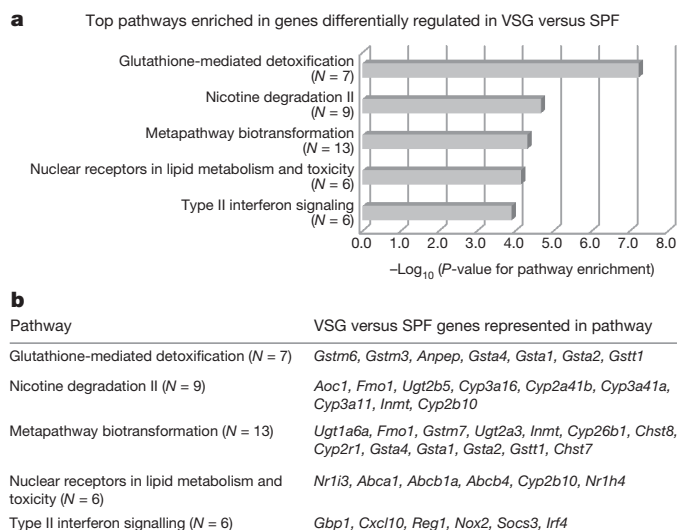


Figure 1 | Unbiased pathway analysis in VSG ileum. **a**, Based on a pathway analysis of genes differentially regulated (fold change ≥ 1.5) in the terminal ileum of VSG mice relative to sham-operated, pair-fed (SPF) controls, the following top five pathways were significantly enriched as follows: glutathione mediated detoxification ($P = 6.00 \times 10^{-8}$), nicotine degradation ($P = 5.64 \times 10^{-4}$), metapathway biotransformation ($P = 4.81 \times 10^{-5}$), nuclear receptors in lipid metabolism and toxicity ($P = 7.18 \times 10^{-5}$), and type II interferon signalling ($P = 1.25 \times 10^{-4}$). *N*, number of genes differentially regulated in each of the five pathways; *n* = 5 mice per group. **b**, Genes enriched in the five pathways identified in **a**.

change ≥ 1.5) in VSG ($P = 7.18 \times 10^{-5}$). Intriguingly, among the other top pathways, several point towards an altered gut microbiota. Specifically, the glutathione pathway ($P = 6.00 \times 10^{-8}$), which plays important roles in nutrient metabolism and antioxidant defence, is known to be significantly altered by the presence/absence of a microbiota—and this is associated with changes in bile acid composition^{11,12}. Likewise, many of the biotransformations ($P = 4.81 \times 10^{-5}$) that are increased probably reflect an altered microbiota, induced for detoxification of metabolites such as secondary bile acids. Finally, changes in gut microbial communities have been extensively linked to altered host immune response including interferon-signalling^{13,14} ($P = 1.25 \times 10^{-4}$).

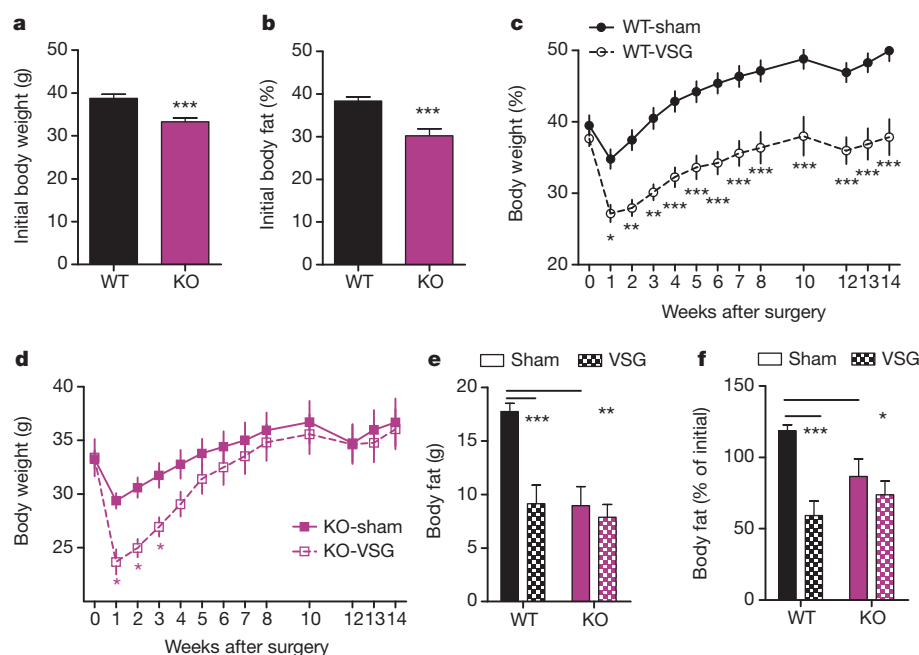


Figure 2 | FXR contributes to the maintenance of weight loss following VSG. **a, b**, Both WT and FXR KO mice weighed more than 30 g (**a**) and carried more than 30% of their weight as fat (**b**) before the surgery. **c, d**, WT-VSG mice lose weight and maintain this weight loss, relative to WT-sham controls (**c**), whereas KO-VSG mice recover the initial weight loss within 5 weeks after surgery, relative to KO-sham controls (**d**). **e**, 11 weeks after surgery, WT-VSG mice carry half the body fat of WT-sham mice whereas the body fat of KO-VSG and KO-sham mice is equivalent. **f**, 11 weeks after surgery, WT-VSG mice have lost 41% of their pre-surgical body fat, whereas KO-VSG mice exhibit no significant fat loss. Data are shown as mean \pm s.e. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. For panels **a–d**, *n* = 12 WT-sham, 8 WT-VSG, 9 KO-sham, 8 KO-VSG. For panels **e** and **f**, *n* = 12 WT-sham, 9 WT-VSG, 9 KO-sham, 10 KO-VSG.

FXR, VSG and body composition

Because we and others have consistently observed increased circulating total bile acids^{9,10,15,16} and increased FXR signalling^{9,17} following various bariatric procedures including VSG, and because *Fxr* was one of the regulated genes identified in our pathway analysis, we hypothesized that FXR-signalling contributes to the metabolic benefits of VSG. To test this, we first generated whole-body FXR knockout mice (KO), and their wild-type littermate controls (WT; Methods). Because KO mice are somewhat resistant to diet-induced obesity¹⁸, we maintained these mice on a 60% high-fat diet (HFD) for 10 weeks before surgery, such that both WT and KO mice became obese. Both genotypes carried more than 30% of their body weight as fat before surgery (Fig. 2a, b); our previous data indicate that this level of adiposity provides more-than-sufficient opportunity for VSG-induced body weight and fat loss^{3,19}.

VSG or sham surgery was performed in these obese male WT and KO mice. In the first week after surgery, WT and KO mice lose a comparable amount of weight relative to their sham-operated controls. However, whereas WT-VSG mice maintained this relative weight loss for the duration of the experiment (Fig. 2c, repeated measures analysis of variance (RM ANOVA) with Tukey post-hoc tests, $P < 0.001$), the KO-VSG animals recovered it. Within 5 weeks, the body weight of KO-VSG animals was no longer different from that of sham-operated controls (Fig. 2d, RM ANOVA with Tukey post hoc, $P < 0.001$). Moreover, when we measured body composition at 11 weeks after the operation, WT-VSG mice had half the body fat of sham-operated WT controls, while the body fat of KO-VSG and KO-sham operated mice was equivalent (Fig. 2e, f, two-way ANOVA with Tukey post hoc, $P < 0.01$).

We note that KO-sham mice continue to gain weight at a slower rate than WT-sham mice following surgery; this needs to be taken into consideration. However, the failure of FXR-VSG mice to sustain body weight and body fat loss is unlikely to be a product of KO mice being relatively resistant to diet-induced obesity, for two reasons. First, we have observed potent reductions in body fat after VSG in mice that have as little as an average 7 g of body fat^{3,19}, whereas the FXR-KO mice in this study had an average of 9 g. Second, we executed post-hoc analyses, selecting the heaviest KO and the lightest WT mice to create a subset with comparable body weights and composition before the surgery. Importantly, in this subset WT-VSG mice maintain an 11.3 g decrement in body weight and a 10.2 g decrement in body fat compared to WT-sham, whereas KO-VSG mice recover to match the body weight and body fat of KO-sham controls within 4–5 weeks. Further, KO-VSG mice are

significantly heavier than WT-VSG mice by 8 weeks after surgery (Extended Data Fig. 1a–c, two-way ANOVA with Tukey post hoc, $P < 0.001$) and maintain a greater fat mass than WT-VSG mice at 11 weeks after surgery (Extended Data Fig. 1d, e, two-way ANOVA with Tukey post hoc, $P < 0.05$).

FXR, VSG and feeding behaviour

To understand how KO-VSG animals were able to recover this body weight, we measured weekly HFD intake in these singly housed mice. Both WT and KO mice consumed fewer calories during the first week following VSG as compared to sham surgery. However, whereas WT-VSG mice maintained this lower food intake for up to 3 weeks (Fig. 3a, RM ANOVA with Tukey post hoc, $P < 0.001$), KO-VSG mice were only hypophagic during the first week (Fig. 3c, RM ANOVA with Tukey post hoc, $P < 0.001$). Importantly, and in contrast to what occurs when animals are calorically-restricted in the absence of surgery²⁰, this caloric deficit was not recovered at later time points in WT-VSG mice, allowing for the maintenance of reduced body weight and body fat over the course of the study (Fig. 3a, b). On the other hand, beginning in post-operation week 4, KO-VSG mice actually consumed more food than sham-operated controls (Fig. 3c, d, Tukey post hoc $P < 0.05$), suggesting that FXR signalling is necessary for the repression of rebound hyperphagia following caloric restriction initially achieved by VSG. This result is also consistent with our growing understanding that VSG is a metabolic, rather than purely restrictive, procedure⁶, as both genotypes are left with a similarly small stomach after VSG.

Eight weeks after surgery, WT-VSG mice had consumed 15% fewer calories than WT-sham mice, but cumulative food intake by KO-VSG mice was equivalent to sham-operated controls (Fig. 3e, two-way ANOVA with Tukey post hoc, $P < 0.01$). Further, when we allowed the mice to choose among three pure macronutrient diets, WT-VSG mice had a significantly reduced preference for fat compared to WT-sham. Unexpectedly, KO-sham mice already exhibit a substantial reduction in the preference for fat compared to WT-sham. This was not further reduced by VSG (Fig. 3f, g, two-way ANOVA with Tukey post hoc, $P < 0.05$). However, we cannot exclude the possibility that the lack of an effect in the KO mice represents a physiological 'floor' that cannot be further reduced by VSG.

FXR, VSG and glucose tolerance

To investigate whether FXR-signalling contributes to improvements in glucose tolerance observed following VSG, we challenged fasted mice with an intraperitoneal (i.p.) injection of dextrose at a dose of 1 g per kg·body weight. VSG was associated with a 20% decrease in fasting

blood glucose in WT mice, whereas it was associated with a 24% increase in fasting blood glucose in KO mice (Fig. 4a, two-way ANOVA with Tukey post hoc, $P < 0.001$). WT-VSG mice exhibited a substantial improvement in the ability to clear the i.p. injection of glucose, reflected as a 35% reduction in the area under the curve (AUC) relative to sham-operated controls. In contrast, KO-VSG and sham-operated mice exhibited no differences in glucose tolerance (Fig. 4b, c, two-way ANOVA with Tukey post hoc, $P < 0.01$; t -test, $P < 0.001$). Although the present findings do not address the question of whether the improvements observed in WT-VSG relative to WT-sham mice are independent of weight loss, we note that KO-VSG and WT-VSG mice had equivalent body weights at the time of this glucose tolerance test. Despite this, when the glucose excursions of WT-VSG and KO-VSG mice are compared directly, KO-VSG mice exhibit significantly impaired glucose clearance at both 30 and 60 min (two-way RM ANOVA with Tukey post hoc, $P < 0.05$, Extended Data Fig. 2).

FXR, VSG and gut microbial communities

Gut microbial communities are altered in obese individuals²¹ and respond to changes in caloric intake and macronutrient content of the diet²². Recent evidence implicates changes in gut microbial communities, both in humans and mice, as potential contributors to the benefits of weight-loss surgery^{23–26}. Transplantation of the gut microbiota from obese mice into germ-free mice is associated with weight gain^{27,28}, whereas transplantation of gut microbiota from RYGB mice is associated with reduced weight gain compared to transplantation with sham microbiota²⁶.

Because bariatric surgery affects the prevalence of many microbial species, it has been difficult to ascertain which specific changes might functionally contribute to its benefits. The present study provides an opportunity to address this question, by comparing the interacting effects of FXR deficiency and VSG on gut microbial ecology. Specifically, because FXR KO mice are refractory to the metabolic benefits of VSG, we infer that any changes in gut microbial communities that occur after VSG irrespective of genotype are not sufficient (but may nonetheless be necessary) to recapitulate its benefits. On the other hand, changes that result from an interaction between genotype and surgery are more likely to be critical to the potent effect of VSG. Pertinent to this, bile acids can modulate the abundance and composition of gut microbes, both directly, and indirectly by activation of FXR-signalling^{29,30}. Conversely, the gut microbiota can modify bile acid composition by microbial enzymatic activities, thereby altering both bile acid homeostasis and FXR signalling^{31,32}.

Mice were killed 14 weeks following surgery, and we extracted genomic DNA from caecal contents and performed pyrosequencing of the

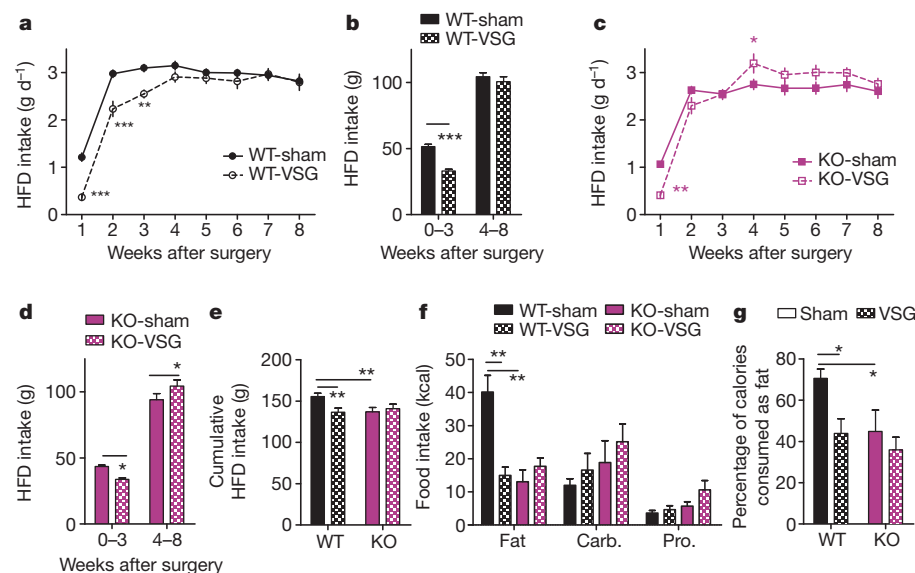


Figure 3 | FXR alters feeding behaviour after VSG. **a, b**, WT mice eat significantly less during the first 3 weeks after VSG versus sham surgery; this is not recovered at later time points. **c, d**, KO mice eat significantly less during the first week after VSG versus sham surgery; beginning at week 4, KO-VSG mice compensate for this initial caloric deficit by consuming more calories than sham-operated controls. **e**, By 8 weeks after surgery, cumulative food intake by WT-VSG mice was 15% less than WT-sham mice, whereas food intake by KO-VSG and sham-operated mice was equivalent. **f**, When given the choice among three pure macronutrient diets, WT-VSG mice exhibit a blunted preference for dietary fat relative to carbohydrates (Carb.) and protein (Pro.). **g**, Among WT mice, VSG reduces the preference for dietary fat. Moreover, loss of FXR is associated with lack of preference for dietary fat that is not further altered by the surgery. Data are shown as mean \pm s.e. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. For panels **a–e**, $n = 13$ WT-sham, 10 WT-VSG, 9 KO-sham, 10 KO-VSG. For panels **f, g**, $n = 12$ WT-sham, 6 WT-VSG, 8 KO-sham, 7 KO-VSG.

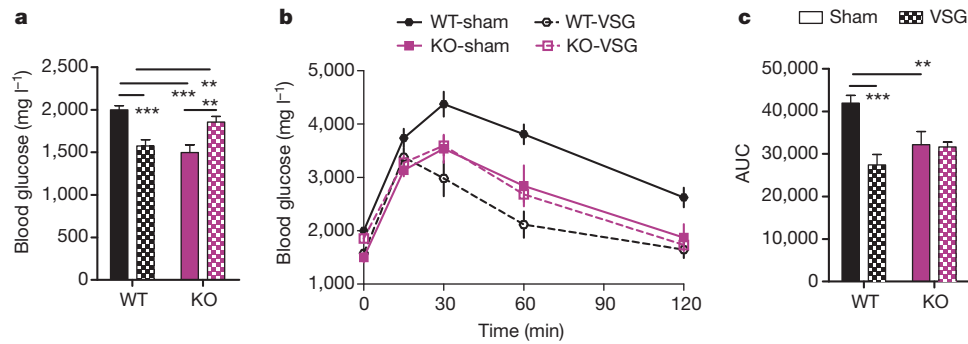


Figure 4 | FXR contributes to the improvement in glucose tolerance observed following VSG. **a**, Among WT mice, fasting blood glucose is reduced following VSG, whereas among KO mice it is increased. **b**, **c**, Among WT mice, VSG improves the excursion of blood glucose following an i.p. bolus of 1 g kg⁻¹

dextrose; among KO mice, the glucose tolerance of VSG and sham-operated mice is equivalent. AUC, area under the curve. Data are shown as mean \pm s.e. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. $n = 13$ WT-sham, 10 WT-VSG, 9 KO-sham, 10 KO-VSG.

16S ribosomal RNA gene. Because all the mice in this study were maintained on the same purified HFD, and because the average caloric intake by WT-VSG mice was only reduced in the first 3 weeks of this 14-week study (Fig. 3), any observed differences are unlikely to be secondary to differing caloric intake or macronutrient content of the diet. Moreover, because WT-VSG, KO-sham and KO-VSG mice had equivalent body weights at death, any observed differences among these groups are unlikely to be secondary to differing body weight or composition.

Weighted UniFrac analysis (sensitive to abundances of taxa)³³ showed separation of WT-VSG samples from WT-sham samples along the first principal component, indicating that VSG had a stronger effect on microbiota composition in WT mice than in KO mice (Fig. 5a, Extended Data Fig. 3, two-way ANOVA with Tukey post hoc). As shown in Fig. 5b, c, the separation of samples in the principal components analysis plot reflects differences in the Bacteroidetes and Firmicutes levels. In light of these differences, we further identified specific taxonomic changes that result from an interaction between genotype and surgery, as these may represent microbial groups that are critical to the potent effects of bariatric surgery—though future studies involving colonization of germ-free mice would be required for definitive evidence supporting a functional involvement.

The relative abundance of several bacteria previously identified as important modulators of systemic metabolism was altered differently by VSG according to genotype. Specifically, the relative abundance of *Bacteroides* was substantially reduced in WT-VSG mice relative to WT-sham controls, but did not vary with surgery among KO mice (Fig. 5d, two-way ANOVA with Tukey post hoc, $P < 0.05$). A recent study investigating the therapeutic potential of probiotic treatment in obese women found that decreasing prevalence of *Bacteroides* was associated with decreasing adiposity and improved glucose control³⁴. Likewise, in the present study, decreasing abundance of *Bacteroides* was associated with decreased weight gain, fat gain (Extended Data Fig. 4), and decreased AUC in a glucose tolerance test—suggesting a functional association that depends on FXR.

An uncharacterized genus in the Porphyromonadaceae family, a bacterial group negatively correlated with type-1 diabetes³⁵, and decreased in HFD-fed mice²², was substantially increased in WT-VSG mice relative to WT-sham controls. In contrast, this group was reduced in KO-VSG mice relative to their sham-operated controls (Fig. 5e, two-way ANOVA with Tukey post hoc, $P < 0.05$).

Finally, the relative abundance of *Roseburia* was 12-fold higher in WT-VSG caeca compared to WT-sham, but the abundance of this microbe in both KO groups resembled the WT-sham group (Fig. 5f, two-way ANOVA with Tukey post hoc, $P < 0.05$). Importantly, and consistent with our findings, two recent gut metagenome analyses identified decreased abundance of *Roseburia* as highly discriminant for type-2 diabetes mellitus, in both a European and a Chinese population^{36,37}.

Likewise, a study investigating the therapeutic effect of a prebiotic treatment in rodents found that increased abundance of *Roseburia* was associated with weight loss and improved glucose tolerance, independent of food intake³⁸. In the present study, the abundance of *Roseburia* was negatively correlated with AUC in the glucose tolerance test (Extended Data Fig. 4), and this was due primarily to the relationship within the WT-VSG group ($R^2 = 0.75$, $P < 0.05$, Fig. 5g), suggesting an FXR-dependent functional association.

In addition, we observed genotype-independent changes in the relative abundance of the genera *Lactobacillus* and *Lactococcus* and of Enterobacteriaceae (Extended Data Fig. 5). Consistent with this, an increase in *Escherichia coli* and other enterobacteria has likewise been reported after RYGB^{23,24,26}. Such changes may indeed contribute to the benefits of VSG and other bariatric procedures, but the present data suggest they are not sufficient to elicit improvements in energy balance and glucose homeostasis.

Changes in gut microbiota may influence host metabolism in part due to the ability of the microbiota to attain colonic dietary conversion, resulting mostly in the production of fatty acid end products. Pertinent to this, a wide range of data now links the production of short-chain fatty acids (SCFAs) to various metabolic outcomes^{36,37,39,40}. To this end, we measured caecal SCFAs and other organic acids produced by the gut microbiota. There were no differences in the total abundance of SCFAs among the four groups (Extended Data Table 1). The relative abundances of butyrate and propionate (Extended Data Fig. 6a, b), but not acetate (Extended Data Fig. 6c) were altered by VSG and this did not differ between genotypes (two-way ANOVA with Tukey post hoc, $P < 0.05$). The resulting decrease in the acetate:butyrate ratio (Extended Data Fig. 6d, two-way ANOVA with Tukey post hoc, $P < 0.001$) possibly indicates that acetate is more efficiently converted to butyrate following VSG. Consistent with the increase in *Lactobacillus* and *Lactococcus*, we also observed an increase in lactate following VSG (Extended Data Table 1). Again, although such changes may indeed contribute to the benefits of VSG and other bariatric procedures, the present data suggest they are not sufficient to elicit metabolic improvements.

Discussion

Though a number of local physiological and environmental variables may be altered by VSG, here we have identified a critical role for altered bile acid signalling. Taken together, the present results demonstrate that a functional FXR pathway is necessary for sustained weight loss, suppression of rebound hyperphagia and improved glucose control following VSG. Importantly, KO mice have an equivalently small stomach, yet they do not demonstrate the same improvements as WT mice. Thus our findings directly contradict the common assumption that mechanical restriction, resulting from the reduced stomach size, is sufficient to mediate the therapeutic effects of VSG.

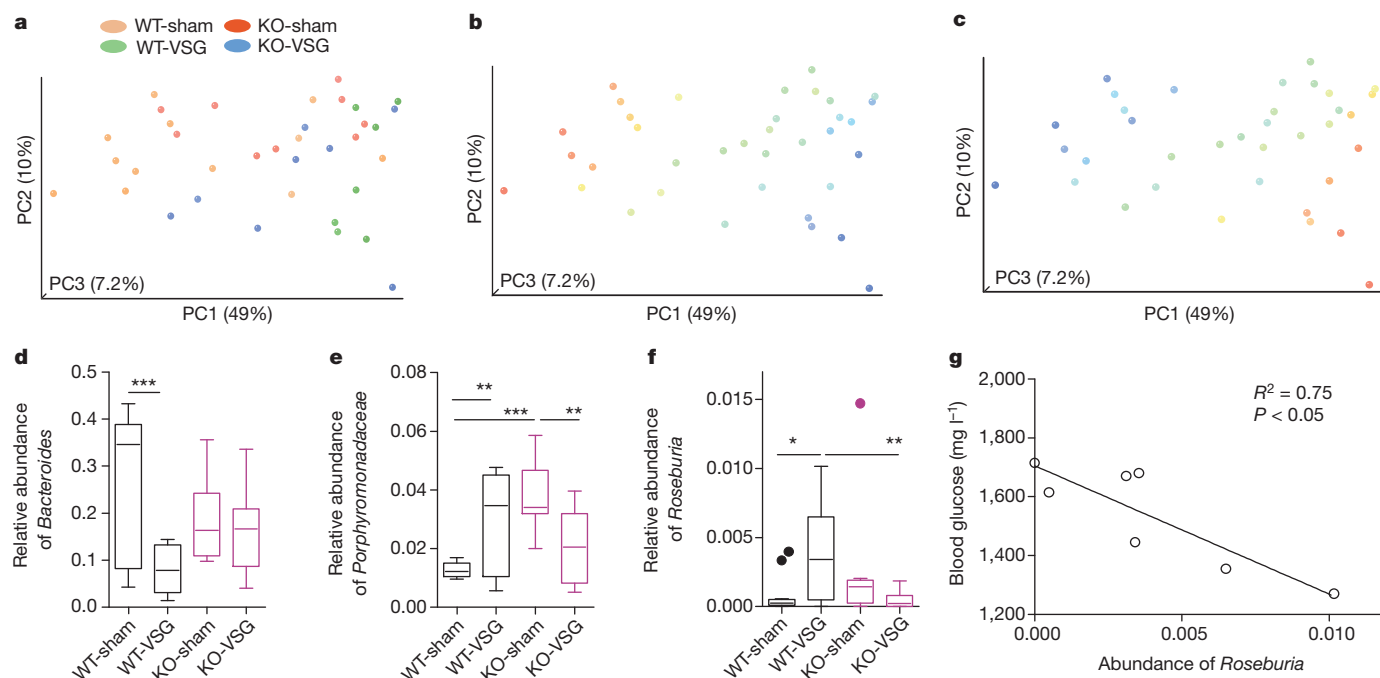


Figure 5 | VSG and FXR alter the composition of caecal microbial communities. **a**, Principal coordinates analysis plot of weighted UniFrac distances. Each dot represents a caecal community with the various surgical/genotype groups indicated by colour. The percentage of variation explained by each principal coordinate (PC1, PC2, PC3) is shown in parentheses. **b**, **c**, Abundance gradient of Bacteroidetes (**b**) and Firmicutes (**c**). Each dot is in the same position as it was in **a** but the colour now indicates the relative abundance (blue indicating low abundance to red indicating high abundance) of Bacteroidetes and Firmicutes respectively. **d**, Relative abundance of *Bacteroides* was reduced in WT-VSG compared to WT-sham operated mice,

but did not differ in KO mice. **e**, Relative abundance of one genus in the Porphyromonadaceae family was increased in WT-VSG compared to WT-sham operated mice, but was reduced in KO-VSG mice relative to KO-sham operated controls. **f**, Relative abundance of *Roseburia* was increased in WT-VSG compared to WT-sham operated mice, but did not differ in KO mice; individual data points show outliers. **g**, Among WT-VSG mice, the relative abundance of *Roseburia* was tightly correlated with fasting blood glucose. Data are presented as Tukey box-plots. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. $n = 12$ WT-sham, 7 WT-VSG, 9 KO-sham, 8 KO-VSG.

In agreement with this, FXR-signalling is implicated in the regulation of lipid and glucose metabolism^{41–44}. Initial reports emphasized a protective role, concluding that FXR signalling is necessary for normal glucose homeostasis. FXR KO mice exhibit impaired glucose and insulin tolerance when maintained on a chow diet^{41,42}, despite modestly elevated plasma bile acids. In apparent contrast, however, they are resistant to HFD-induced obesity and glucose intolerance¹⁸. Intriguingly, two studies investigating the effect of the synthetic FXR agonist GW4064 on the development of HFD-induced obesity and glucose intolerance^{45,46} yielded opposite results. Whereas Watanabe *et al.*⁴⁶ report greater sensitivity to diet-induced metabolic dysfunction, Ma *et al.*⁴⁵ report resistance to diet-induced obesity following treatment with GW4064. A major difference between the two studies was route of administration. Watanabe *et al.* delivered the drug orally via the diet, whereas Ma *et al.* delivered it intraperitoneally. FXR is expressed in multiple organs important for the regulation of metabolism, including liver, adipose tissue and intestine, and its activation in these compartments may elicit opposing effects on metabolism, perhaps contributing to this discrepancy. Relevant to this, bile acids may also affect metabolism by signalling via the G-protein-coupled receptor TGR5, raising the possibility of cross-talk between these systems⁴⁷. Finally, we note these contradictory phenotypes have likewise been observed in the complex regulation of gene expression by other members of the nuclear receptor family. The PPAR γ haploinsufficient mouse, for example, is more insulin sensitive despite the fact that PPAR γ agonists are well-known for their insulin-sensitizing effects⁴⁸. Like PPAR γ , FXR and other nuclear receptors may either activate or repress the transcription of target genes, depending on the presence of co-activator and co-repressor complexes. Therefore, loss-of-FXR function might paradoxically be associated with a relative increased expression in a subset of FXR-regulated genes, while still making the animal

resistant to treatments that activate FXR. This may explain the present conundrum that FXR KO mice are simultaneously resistant to some of the deleterious metabolic effects of consuming a HFD and also relatively unresponsive to the metabolic benefits of VSG.

Bariatric surgical procedures, such as VSG, are currently the most effective therapy for the treatment of obesity and are moreover associated with substantial improvements in co-morbidities, including type-2 diabetes mellitus. The underlying molecular mechanisms contributing to these benefits remain largely undetermined, despite offering tremendous potential to reveal new targets for therapeutic intervention. Progress towards this goal has probably been hindered by a widespread belief that surgery exerts its effects simply by making it physically difficult to consume or absorb calories. The current work provides an alternative framework—that is, it identifies FXR-signalling as a molecular mechanism necessary for both key changes to gut microbial communities and many associated metabolic benefits of VSG. These findings provide a new understanding of the mechanisms that underlie the benefits of bariatric surgery and suggest new targets for developing less-invasive therapeutic interventions.

METHODS SUMMARY

RNA-seq and unbiased pathway analysis. These were performed using RNA collected from the distal small intestine of VSG and sham-operated, pair-fed male (SPF) mice that were the subjects of a previously published study⁹.

Animals. Singly-housed FXR KO mice and their WT littermates⁴⁹ were allowed *ad libitum* access to a high-fat diet (HFD) and water unless otherwise noted. VSG or sham surgery was performed as described elsewhere³.

Microbiota. Caecal microbiota composition was analysed by pyrosequencing of the V1–V2 region from the bacterial 16S rRNA gene. PCR amplicons were generated from caecal DNA. Sequence data were analysed by using the QIIME software package (version 1.5.0)⁵⁰.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 12 July 2013; accepted 6 February 2014.

Published online 26 March 2014.

- Pories, W. J. Bariatric surgery: risks and rewards. *J. Clin. Endocrinol. Metab.* **93**, S89–S96 (2008).
- Schauer, P. R. *et al.* Bariatric surgery versus intensive medical therapy in obese patients with diabetes. *N. Engl. J. Med.* **366**, 1567–1576 (2012).
- Wilson-Pérez, H. E. *et al.* Vertical sleeve gastrectomy is effective in two genetic mouse models of glucagon-like peptide-1 receptor deficiency. *Diabetes* **62**, 2380–2385 (2013).
- Chambers, A. P. *et al.* Weight-independent changes in blood glucose homeostasis after gastric bypass or vertical sleeve gastrectomy in rats. *Gastroenterology* **141**, 950–958 (2011).
- Karamanakos, S. N., Vagenas, K., Kalfarentzos, F. & Alexandrides, T. K. Weight loss, appetite suppression, and changes in fasting and postprandial ghrelin and peptide-YY levels after Roux-en-Y gastric bypass and sleeve gastrectomy: a prospective, double blind study. *Ann. Surg.* **247**, 401–407 (2008).
- Stefater, M. A., Wilson-Pérez, H. E., Chambers, A. P., Sandoval, D. A. & Seeley, R. J. All bariatric surgeries are not created equal: insights from mechanistic comparisons. *Endocr. Rev.* **33**, 595–622 (2012).
- Jackness, C. *et al.* Very low-calorie diet mimics the early beneficial effect of Roux-en-Y gastric bypass on insulin sensitivity and β -cell function in type 2 diabetic patients. *Diabetes* **62**, 3027–3032 (2013).
- Bradley, D. *et al.* Gastric bypass and banding equally improve insulin sensitivity and β cell function. *J. Clin. Invest.* **122**, 4667–4674 (2012).
- Myronovych, A. *et al.* Vertical sleeve gastrectomy reduces hepatic steatosis while increasing serum bile acids in a weight-loss-independent manner. *Obesity* **22**, 390–400 (2014).
- Patti, M. E. *et al.* Serum bile acids are higher in humans with prior gastric bypass: potential contribution to improved glucose and lipid metabolism. *Obesity* **17**, 1671–1677 (2009).
- Larsson, E. *et al.* Analysis of gut microbial regulation of host gene expression along the length of the gut and regulation of gut microbial ecology through MyD88. *Gut* **61**, 1124–1131 (2012).
- Lhoste, E. F. *et al.* The human colonic microflora influences the alterations of xenobiotic-metabolizing enzymes by catechins in male F344 rats. *Food Chem. Toxicol.* **41**, 695–702 (2003).
- Brodziak, F., Meharg, C., Blaut, M. & Loh, G. Differences in mucosal gene expression in the colon of two inbred mouse strains after colonization with commensal gut bacteria. *PLoS ONE* **8**, e72317 (2013).
- Vijay-Kumar, M. *et al.* Metabolic syndrome and altered gut microbiota in mice lacking Toll-like receptor 5. *Science* **328**, 228–231 (2010).
- Kohli, R. *et al.* Weight loss induced by Roux-en-Y gastric bypass but not laparoscopic adjustable gastric banding increases circulating bile acids. *J. Clin. Endocrinol. Metab.* **98**, E708–E712 (2013).
- Cummings, B. P. *et al.* Vertical sleeve gastrectomy improves glucose and lipid metabolism and delays diabetes onset in UCD-T2DM rats. *Endocrinology* **153**, 3620–3632 (2012).
- Gerhard, G. S. *et al.* A role for fibroblast growth factor 19 and bile acids in diabetes remission after Roux-en-Y gastric bypass. *Diabetes Care* **36**, 1859–1864 (2013).
- Prawitt, J. *et al.* Farnesoid X receptor deficiency improves glucose homeostasis in mouse models of obesity. *Diabetes* **60**, 1861–1871 (2011).
- Chambers, A. P. *et al.* The effects of vertical sleeve gastrectomy in rodents are ghrelin independent. *Gastroenterology* **144**, 50–52.e5 (2013).
- Ryan, K. K., Woods, S. C. & Seeley, R. J. Central nervous system mechanisms linking the consumption of palatable high-fat diets to the defense of greater adiposity. *Cell Metab.* **15**, 137–149 (2012).
- Ley, R. E. *et al.* Obesity alters gut microbial ecology. *Proc. Natl Acad. Sci. USA* **102**, 11070–11075 (2005).
- Parks, B. W. *et al.* Genetic control of obesity and gut microbiota composition in response to high-fat, high-sucrose diet in mice. *Cell Metab.* **17**, 141–152 (2013).
- Zhang, H. *et al.* Human gut microbiota in obesity and after gastric bypass. *Proc. Natl Acad. Sci. USA* **106**, 2365–2370 (2009).
- Furet, J.-P. *et al.* Differential adaptation of human gut microbiota to bariatric surgery-induced weight loss: links with metabolic and low-grade inflammation markers. *Diabetes* **59**, 3049–3057 (2010).
- Graessler, J. *et al.* Metagenomic sequencing of the human gut microbiome before and after bariatric surgery in obese patients with type 2 diabetes: correlation with inflammatory and metabolic parameters. *Pharmacogenom. J.* **13**, 514–522 (2012).
- Liou, A. P. *et al.* Conserved shifts in the gut microbiota due to gastric bypass reduce host weight and adiposity. *Sci. Transl. Med.* **5**, 178ra41 (2013).
- Turnbaugh, P. J. *et al.* An obesity-associated gut microbiome with increased capacity for energy harvest. *Nature* **444**, 1027–1031 (2006).
- Turnbaugh, P. J., Bäckhed, F., Fulton, L. & Gordon, J. I. Diet-induced obesity is linked to marked but reversible alterations in the mouse distal gut microbiome. *Cell Host Microbe* **3**, 213–223 (2008).
- Merritt, M. E. & Donaldson, J. R. Effect of bile salts on the DNA and membrane integrity of enteric bacteria. *J. Med. Microbiol.* **58**, 1533–1541 (2009).
- Islam, K. B. M. S. *et al.* Bile acid is a host factor that regulates the composition of the cecal microbiota in rats. *Gastroenterology* **141**, 1773–1781 (2011).
- Sayin, S. I. *et al.* Gut microbiota regulates bile acid metabolism by reducing the levels of tauro- β -muricholic acid, a naturally occurring FXR antagonist. *Cell Metab.* **17**, 225–235 (2013).
- Swann, J. R. *et al.* Systemic gut microbial modulation of bile acid metabolism in host tissue compartments. *Proc. Natl Acad. Sci. USA* **108** (Suppl. 1), 4523–4530 (2011).
- Lozupone, C. & Knight, R. UniFrac: a new phylogenetic method for comparing microbial communities. *Appl. Environ. Microbiol.* **71**, 8228–8235 (2005).
- Dewulf, E. M. *et al.* Insight into the prebiotic concept: lessons from an exploratory, double blind intervention study with inulin-type fructans in obese women. *Gut* **62**, 1112–1121 (2013).
- Wen, L. *et al.* Innate immunity and intestinal microbiota in the development of Type 1 diabetes. *Nature* **455**, 1109–1113 (2008).
- Qin, J. *et al.* A metagenome-wide association study of gut microbiota in type 2 diabetes. *Nature* **490**, 55–60 (2012).
- Karlsson, F. H. *et al.* Gut metagenome in European women with normal, impaired and diabetic glucose control. *Nature* **498**, 99–103 (2013).
- Neyrinck, A. M. *et al.* Dietary modulation of clostridial cluster XIVa gut bacteria (*Roseburia* spp.) by chitin-glucan fiber improves host metabolic alterations induced by high-fat diet in mice. *J. Nutr. Biochem.* **23**, 51–59 (2012).
- Ridaura, V. K. *et al.* Gut microbiota from twins discordant for obesity modulate metabolism in mice. *Science* **341**, 1241214 (2013).
- Vrieze, A. *et al.* Transfer of intestinal microbiota from lean donors increases insulin sensitivity in individuals with metabolic syndrome. *Gastroenterology* **143**, 913–6.e7 (2012).
- Cariou, B. *et al.* The farnesoid X receptor modulates adiposity and peripheral insulin sensitivity in mice. *J. Biol. Chem.* **281**, 11039–11049 (2006).
- Ma, K., Saha, P. K., Chan, L. & Moore, D. D. Farnesoid X receptor is essential for normal glucose homeostasis. *J. Clin. Invest.* **116**, 1102–1109 (2006).
- Zhang, Y. *et al.* Activation of the nuclear receptor FXR improves hyperglycemia and hyperlipidemia in diabetic mice. *Proc. Natl Acad. Sci. USA* **103**, 1006–1011 (2006).
- Porez, G., Prawitt, J., Gross, B. & Staels, B. Bile acid receptors as targets for the treatment of dyslipidemia and cardiovascular disease. *J. Lipid Res.* **53**, 1723–1737 (2012).
- Ma, Y., Huang, Y., Yan, L., Gao, M. & Liu, D. Synthetic FXR agonist GW4064 prevents diet-induced hepatic steatosis and insulin resistance. *Pharm. Res.* **30**, 1447–1457 (2013).
- Watanabe, M. *et al.* Lowering bile acid pool size with a synthetic farnesoid X receptor (FXR) agonist induces obesity and diabetes through reduced energy expenditure. *J. Biol. Chem.* **286**, 26913–26920 (2011).
- Thomas, C., Pellicciari, R., Pruzanski, M., Auwerx, J. & Schoonjans, K. Targeting bile-acid signalling for metabolic diseases. *Nature Rev. Drug Discov.* **7**, 678–693 (2008).
- Miles, P. D., Barak, Y., He, W., Evans, R. M. & Olefsky, J. M. Improved insulin-sensitivity in mice heterozygous for PPARG deficiency. *J. Clin. Invest.* **105**, 287–292 (2000).
- Sinal, C. J. *et al.* Targeted disruption of the nuclear receptor FXR/BAR impairs bile acid and lipid homeostasis. *Cell* **102**, 731–744 (2000).
- Caporaso, J. G. *et al.* QIIME allows analysis of high-throughput community sequencing data. *Nature Methods* **7**, 335–336 (2010).

Acknowledgements We thank J. Berger, A. Haller, B. Li, E. Orr and M. Toure for technical assistance. This work was supported by grants from the UNIK Food Fitness and Pharma for Health and Disease research programme (C.C.), the Torsten Söderberg and NovoNordisk foundations (F.B.), Ethicon Endo-Surgery (R.K., D.A.S., R.J.S.) and the NIH (DK082173, HL111319 to K.K.R., DK093848 to R.J.S. and the Bioinformatics Core of the Digestive Disease Research Core Center in Cincinnati DK078392).

Author Contributions K.K.R. conceptualized, designed, performed and analysed the experiments and wrote the manuscript. C.C., A.M., H.E.W.-P., D.A.S. and R. Kohli performed experiments and edited the manuscript. R. Karns performed the bioinformatics analysis of the RNA-seq data. V.T. and F.B. designed and performed the microbiota analysis and edited the manuscript. P.K.-D. and F.B. designed and performed the analysis of caecal metabolites and edited the manuscript. R.J.S. conceptualized, designed and analysed the experiments and wrote the manuscript.

Author Information Raw and normalized RNA-seq data have been deposited in the NCBI Gene Expression Omnibus database under accession number GSE53782. Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to R.J.S. (randy.seeley@uc.edu).

METHODS

Protocols. The Institutional Animal Care and Use Committees at the University of Cincinnati and Cincinnati Children's Hospital Medical Center approved all animal protocols.

RNA-seq and unbiased pathway analysis. For RNA-seq followed by unbiased pathway analysis, the distal 1.5 cm of small intestine was collected from 5 VSG and 5 sham-operated, pair-fed male (SPF) C57Bl/6J mice (Jackson Labs, BarHarbour, ME) that were the subjects of a previously published study⁹. See ref. 9 for further methodological detail. RNA-seq libraries prepared with the Illumina TruSeq RNA preparation kit were sequencing on an Illumina Hi-Seq 2000, with single-end 50 bp reads. Analysis was performed in GeneSpring NGS (Agilent Technologies, Santa Clara, CA), where sequences were aligned to the mm9 reference mouse genome and assigned Ensembl annotations. Aligned gene reads that passed quality controls (quality threshold ≥ 30 , zero Ns allowed, no multiply mapping reads) were quantified and used to calculate reads per kilobase per million (RPKM). Raw data were normalized using the DeSeq algorithm where reads were thresholded to a minimum read count of 1. An additional filter required at least 10 reads per entity in each sample of either VSG or SPF samples ($N = 14,591$ entities). Differentially expressed genes with a fold change of ≥ 1.5 between VSG and SPF animals ($N = 882$) were submitted to pathway analysis, which assessed significance of overlap between the 882 differential genes and those genes in pathways curated by Wikipathways, BioCyc, BioPax and Legacy, with a P -value cut-off of 0.05.

Surgery. VSG surgery was performed in DIO mice using isoflurane anaesthesia. The lateral 80% of the stomach was excised leaving a tubular gastric remnant in continuity with the oesophagus superiorly and the pylorus and duodenum inferiorly. The sham procedure involved analogous isolation of the stomach followed by manually applying pressure with blunt forceps along a vertical line between the oesophageal sphincter and the pylorus. Mice consumed liquid diet (Osmolite OneCal) for the first 3 post-operative days, and were re-introduced to HFD on day 3.

Mice. 10–12-week-old male FXR KO mice and their WT littermates, described elsewhere⁴⁹ and bred from founders acquired from the Jackson Labs (Bar Harbour, ME), were singly housed in an AAALAC-approved facility with a 12–12 light–dark cycle, and allowed *ad libitum* access to HFD (D12492, Research Diets, New Brunswick, NJ) and water unless otherwise noted. After 10 weeks of HF-feeding, male mice were randomly divided into body weight and fat-matched groups within genotype, and we performed VSG or sham surgery as above. Based on our power analyses, completed using previous data to determine the expected variance and effect sizes, and based on our historical data regarding post-operative mortality, we performed sham surgery on 13 WT and 9 KO mice and VSG surgery on 14 WT and 15 KO mice. 5 KO-VSG and 4 WT-VSG mice died from complications of surgery and were not included in any of the analyses. Food intake and body weight was monitored weekly. Body composition was measured 1 week before and 11 weeks following surgery by magnetic resonance imaging (Echo Medical Systems, Houston, TX). An additional 1 WT-sham, 2 KO-VSG and 2 WT-VSG mice died (or were found moribund) during the period of metabolic phenotyping that occurred between weeks 8 and 15. For any given analysis, only mice that are represented throughout the entire timecourse were included. In addition, for the macronutrient preference study, data from several individuals were excluded based on pre-established criteria—in this case, that the mice consistently spilled the powdered macronutrient diets on the cage floor making accurate measurement of what was ingested impossible. For the microbiota analyses, one WT-VSG individual was excluded due to a noted investigator error while collecting the caecal samples at sacrifice. Investigators conducting the gene expression and microbiota analyses were blind to the experimental groups.

Macronutrient preference. Food choice was assayed during postoperative week 10 using a macronutrient selection paradigm, in which 3 pure macronutrient diets (Harlan Teklad; TD.02521 (carbohydrate), TD.02522 (fat) and TD02523 (protein)) were presented simultaneously in separate containers for 6 days ($n = 12$ WT-sham, 6 WT-VSG, 8 KO-sham, 7 KO-VSG).

Glucose tolerance tests. Glucose tolerance was assayed during postoperative week 8 in 5-h fasted mice. Blood glucose was measured with AccuCheck meters and strips (Roche, Indianapolis, IN) from the tip of the tail vein in unrestrained animals at baseline, and at regular intervals following an i.p. injection of 1 g kg^{-1} body weight dextrose ($n = 13$ WT-sham, 10 WT-VSG, 9 KO-sham, 10 KO-VSG).

Analysis of gut microbial communities. Whereas much of the microbiota is dead in the faeces, the caecum is an active bioreactor with high metabolic fermentative capacity. Thus we focused our study on the caecal microbiota. Caecal microbiota composition was analysed by pyrosequencing of the V1–V2 region from the bacterial 16S rRNA gene. PCR amplicons were generated from caecal DNA obtained from WT and FXR KO mice that underwent either VSG or sham operations ($n = 12$ WT-sham, 7 WT-VSG, 9 KO-sham, 8 KO-VSG). Sequence data were analysed by using the QIIME software package (version 1.5.0)⁵⁰.

Sample processing, extraction of genomic DNA and amplification for 454 pyrosequencing. Aliquots of caecal material were collected and immediately frozen in liquid nitrogen. These aliquots were stored at -80°C until extracted: approximately half of each aliquot was used for isolation of genomic DNA and the other half was used for extraction of SCFAs and microbial metabolites.

Genomic DNA was isolated from 50–80 mg of caecal content using the repeated bead beating (RBB) method previously described by Salonen *et al.*⁵¹. The V1–V2 region of the 16S rRNA gene was amplified using the 27F and 338R primers fused with 454 Titanium sequencing adapters. 338R primers contained unique error-correcting 12-base barcodes that allow to tag PCR products from different samples⁵². Each sample was amplified in triplicate in a reaction volume of 25 μl containing 1.5 U of FastStart Taq DNA Polymerase (Roche), 0.2 μM of each primer and 10–20 ng of genomic DNA. PCR was carried out under the following conditions: initial denaturation for 3 min at 95°C , followed by 25 cycles of denaturation for 20 s at 95°C , annealing for 20 s at 52°C and elongation for 60 s at 72°C , and a final elongation step for 8 min at 72°C . Triplicates were combined, purified with the NucleoSpin Gel and PCR Clean-up kit (Macherey-Nagel, Germany) and then quantified using the Quant-iT PicoGreen dsDNA kit (Invitrogen, Carlsbad, CA). Purified PCR products were diluted to a concentration of 20 $\text{ng } \mu\text{l}^{-1}$ and pooled in equal amounts. The pooled amplicons were purified again with the Ampure magnetic purification beads (Agencourt, Danvers, MA) to remove short amplification products. Sequencing was performed using 454 GS FLX titanium chemistry at GATC Biotech (Konstanz, Germany).

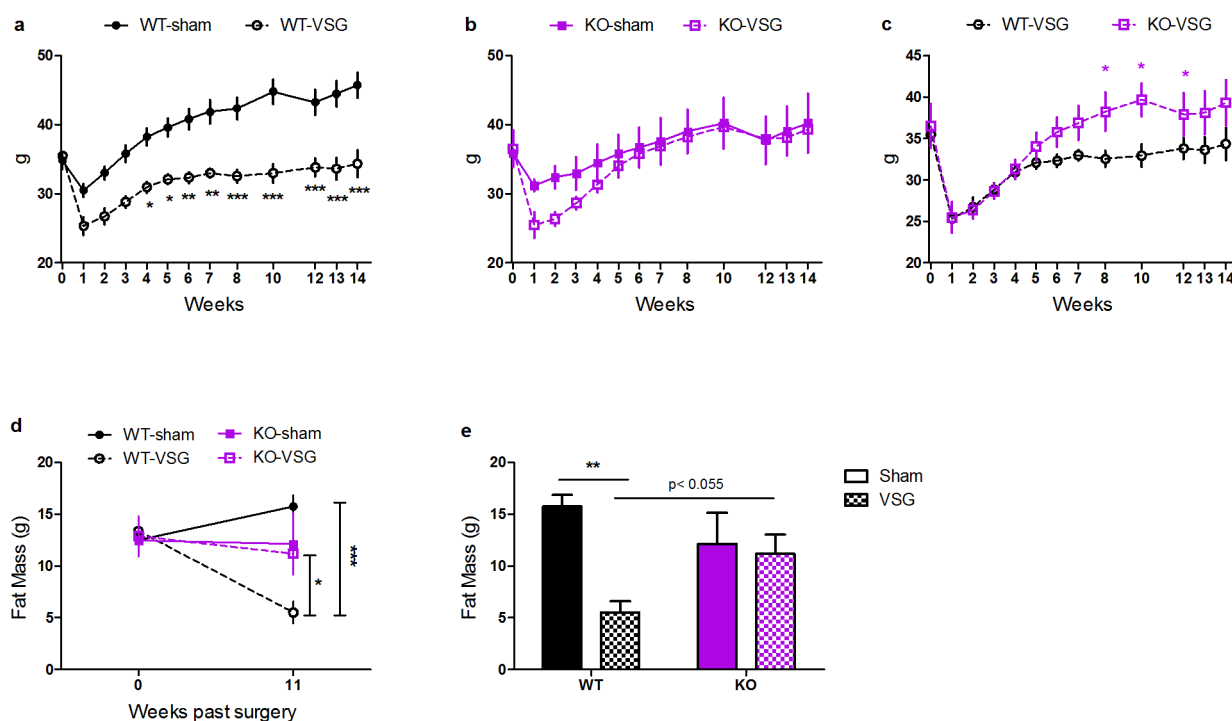
Analysis of 16S rRNA gene sequences. Raw data were quality filtered to remove sequences that were shorter than 200 nucleotides, longer than 1,000 nucleotides, contained primer mismatches, ambiguous bases, uncorrectable barcodes, or homopolymer runs in excess of six bases. Quality filtered reads were trimmed of 454 adaptor and barcode sequences, and were analysed with the software package Quantitative Insights Into Microbial Ecology (QIIME)⁵⁰ (version 1.5.0). A total of 376,965 sequences were obtained for 36 samples; sequences were demultiplexed and an average of 10,471 sequences were attributed to each sample (range: 7,208–12,522 sequences).

Sequences were assigned to operational taxonomic units (OTUs) using UCLUST with a 97% threshold of pairwise identity. The most abundant sequence was picked as representative for each OTU and was given taxonomic assignment using the Ribosomal Database Project (RDP) Classifier⁴. Representative OTUs were aligned using Pynast⁵³ and used to build a phylogenetic tree with FastTree⁵⁴, which was used to estimate β -diversity of samples (weighted UniFrac³³). Sequences were checked for chimerae using ChimeraSlayer and chimeraic sequences were excluded from all downstream analyses. Similarly, singletons and sequences that could not be aligned with Pynast were also excluded.

Extraction of microbial metabolites (SCFAs and other organic acids). GC-MS was used for measurement of organic acids in mouse caecal samples ($n = 12$ WT-sham, 7 WT-VSG, 9 KO-sham, 8 KO-VSG). After extraction with diethyl ether, an aliquot of each sample was derivatized with *N*-tert-butyltrimethylsilyl-*N*-methyltrifluoroacetamide (MTBSTFA; Sigma), and the organic acids were quantified by using a gas chromatograph (Model 7890 A, Agilent Technologies) coupled to a mass spectrometer detector (Model 5975 C, inert XL MSD with triple Axis Detector; Agilent Technologies). 50–80 mg of frozen caecal contents were transferred to glass tubes ($16 \times 125 \text{ mm}$) fitted with a screw cap, and a volume of 100 μl of internal standards stock solution ($[1\text{-}^{13}\text{C}]$ acetate and $[2\text{-}^3\text{H}_6]$ propionate 1 M, $[1\text{-}^{13}\text{C}_4]$ butyrate 0.5 M, $[1\text{-}^{13}\text{C}_4]$ isobutyrate and $[1\text{-}^{13}\text{C}]$ isovalerate 0.1 M, $[1,2\text{-}^{13}\text{C}_2]$ hexanoate, $[1\text{-}^{13}\text{C}]$ lactate and $[1\text{-}^{13}\text{C}_4]$ succinic acid each at 40 mM) was added to the tubes. Prior to extraction samples were freeze-dried at -50°C for 3 h (yield 8–15 mg dry weight). After acidification with 50 μl of 37% HCl, the organic acids were extracted twice in 2 ml of diethyl ether. A 500 μl aliquot of the extracted sample was mixed together with 50 μl of MTBSTFA (Sigma) at room temperature. An aliquot (1 μl) of the resulting derivatized material was injected into a gas chromatograph (Agilent Technologies 7890 A) coupled to a mass spectrometer detector (Agilent Technologies 5975 C). A linear temperature gradient was used: the initial temperature of 65°C was held for 6 min, increased to 260°C ($15^\circ\text{C min}^{-1}$) and then to 280°C for 5 min. The injector and transfer line temperatures were 250°C . Quantitation was completed in selected ion monitoring acquisition mode by comparison to labelled internal standards (valerate was compared to $[1\text{-}^{13}\text{C}]$ isovalerate, heptanoate and octanoate were compared to $[1,2\text{-}^{13}\text{C}_2]$ hexanoate and fumarate was compared to $[1\text{-}^{13}\text{C}_4]$ succinic acid). The m/z ratios of monitored ions were as follows: 117 (acetic acid), 131 (propionic acid), 145 (butyric acid), 146 (isovaleric acid), 159 (isovaleric acid and valeric acid), 173 (hexanoic acid), 187 (heptanoic acid), 201 (octanoic acid), 261 (lactic acid), 287 (fumaric acid), 289 (succinic acid), 121 ($[2\text{-}^3\text{H}_2]$ - and $[1\text{-}^{13}\text{C}]$ acetate), 136 ($[2\text{-}^3\text{H}_5]$ propionate), 146 ($[1\text{-}^{13}\text{C}_4]$ isobutyrate), 149 ($[1\text{-}^{13}\text{C}_4]$ butyrate), 160 ($[1\text{-}^{13}\text{C}]$ isovalerate), 175 ($[1,2\text{-}^{13}\text{C}_2]$ hexanoate), 264 ($[1\text{-}^{13}\text{C}]$ lactate) and 293 ($[1\text{-}^{13}\text{C}_4]$ succinic acid).

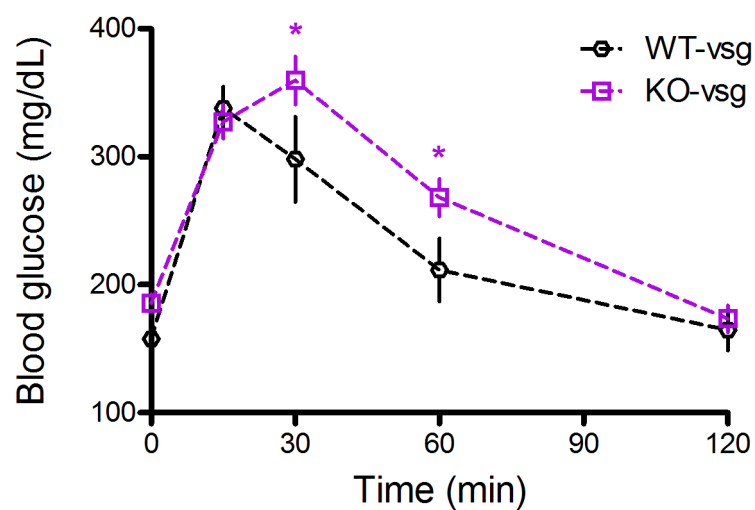
Statistical analyses. Data are presented as mean \pm standard errors or as Tukey box-plots, as noted. Data were analysed using the appropriate ANOVA or repeated measures ANOVA (RM ANOVA), followed by Tukey's post-hoc tests, or by *t*-test as indicated. The data were tested for normality and homogeneity of variance. In cases where the data failed these assumptions, the analyses were also done after rank transformation. In no cases did this alter the statistical significance or interpretation of the findings, so the statistics reported in the figures are from the non-transformed analyses. Data were analysed using Prism (Graph Pad, San Diego CA) or Sigma Stat (SYSTAT, San Jose CA) software with the critical value, α , set at $P < 0.05$.

51. Salonen, A. *et al.* Comparative analysis of fecal DNA extraction methods with phylogenetic microarray: effective recovery of bacterial and archaeal DNA using mechanical cell lysis. *J. Microbiol. Methods* **81**, 127–134 (2010).
52. Hamady, M., Walker, J. J., Harris, J. K., Gold, N. J. & Knight, R. Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods* **5**, 235–237 (2008).
53. Caporaso, J. G. *et al.* PyNAST: a flexible tool for aligning sequences to a template alignment. *Bioinformatics* **26**, 266–267 (2010).
54. Price, M. N., Dehal, P. S. & Arkin, A. P. FastTree: computing large minimum evolution trees with profiles instead of a distance matrix. *Mol. Biol. Evol.* **26**, 1641–1650 (2009).



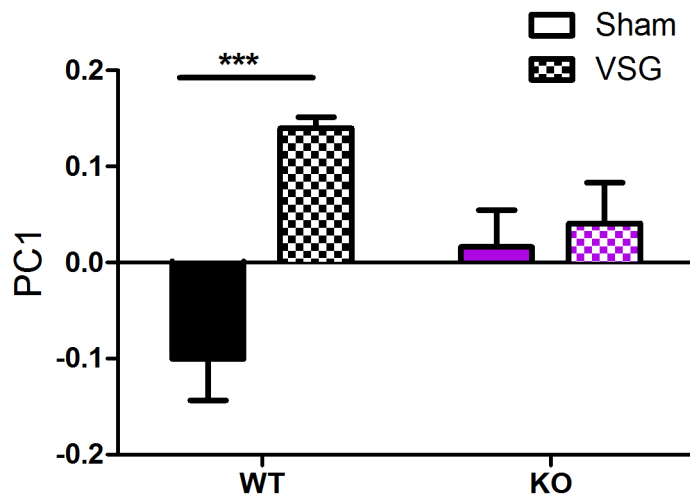
Extended Data Figure 1 | Body weight and body fat in a weight-matched subset of WT and FXR KO mice. These post-hoc analyses include the lightest 9 WT and heaviest 8 FXR-KO mice before surgery, creating 4 well-matched groups. **a–d**, In this subset, WT-VSG lose weight relative to WT-sham controls, and maintain this weight loss for 14 weeks (**a**), whereas KO-VSG mice lose weight initially, but recover to match the weight of KO-sham controls within 4–5 weeks (**b**). Consequently, by week 8 KO-VSG mice were heavier than

WT-VSG mice (**c**). Likewise, these groups were well-matched for pre-surgical body fat (**d**). At 11 weeks after surgery, WT-VSG mice had significantly less body fat compared to both WT-sham controls and KO-VSG mice. KO-sham and KO-VSG mice had equivalent adiposity (**e**). Data are shown as mean \pm s.e. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. For all panels $n = 5$ WT-sham, 4 WT-VSG, 4 KO-sham, 4 KO-VSG.

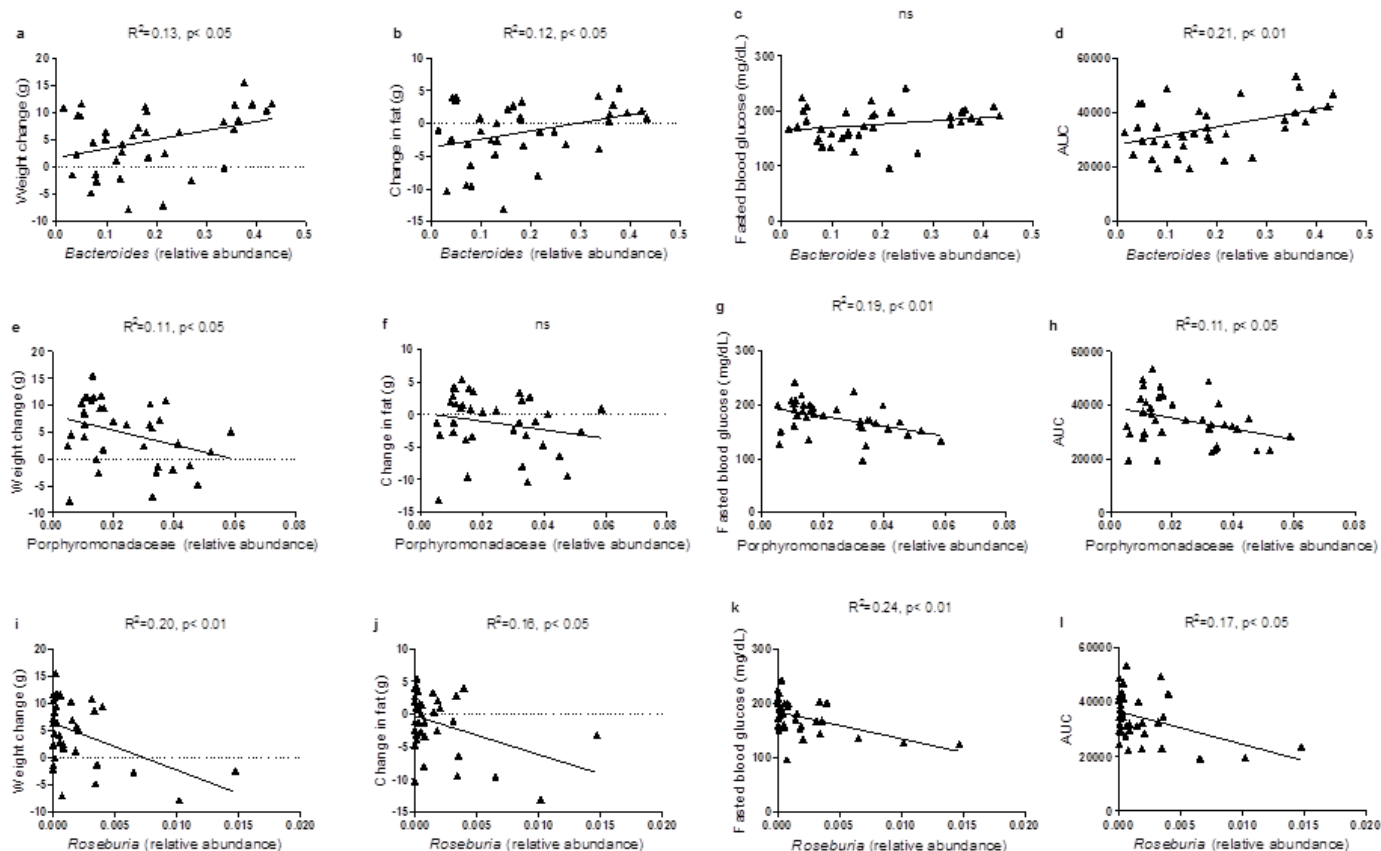


Extended Data Figure 2 | Glucose tolerance in WT-VSG and KO-VSG mice. When the glucose excursion of WT-VSG and KO-VSG mice are compared

directly, KO-VSG mice exhibit significantly impaired glucose clearance at both 30 and 60 min. Data are shown as mean \pm s.e. * $P < 0.05$. $n = 10$ per group.

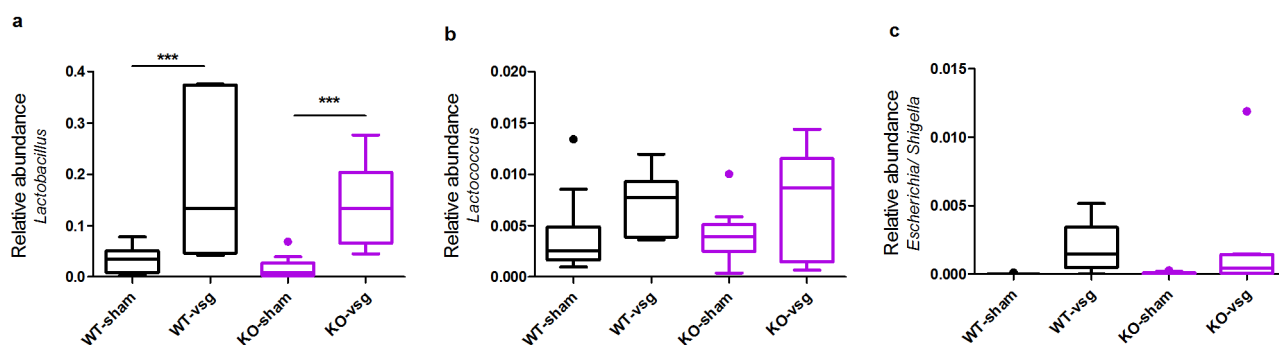


Extended Data Figure 3 | Effect of genotype and VSG on distribution along PC1. Among WT mice, sham and VSG mice separate significantly along PC1. In contrast, among KO mice there is no significant difference between sham and VSG. Data are shown as mean \pm s.e. *** $P < 0.001$. $n = 12$ WT-sham, 7 WT-VSG, 9 KO-sham, 8 KO-VSG.



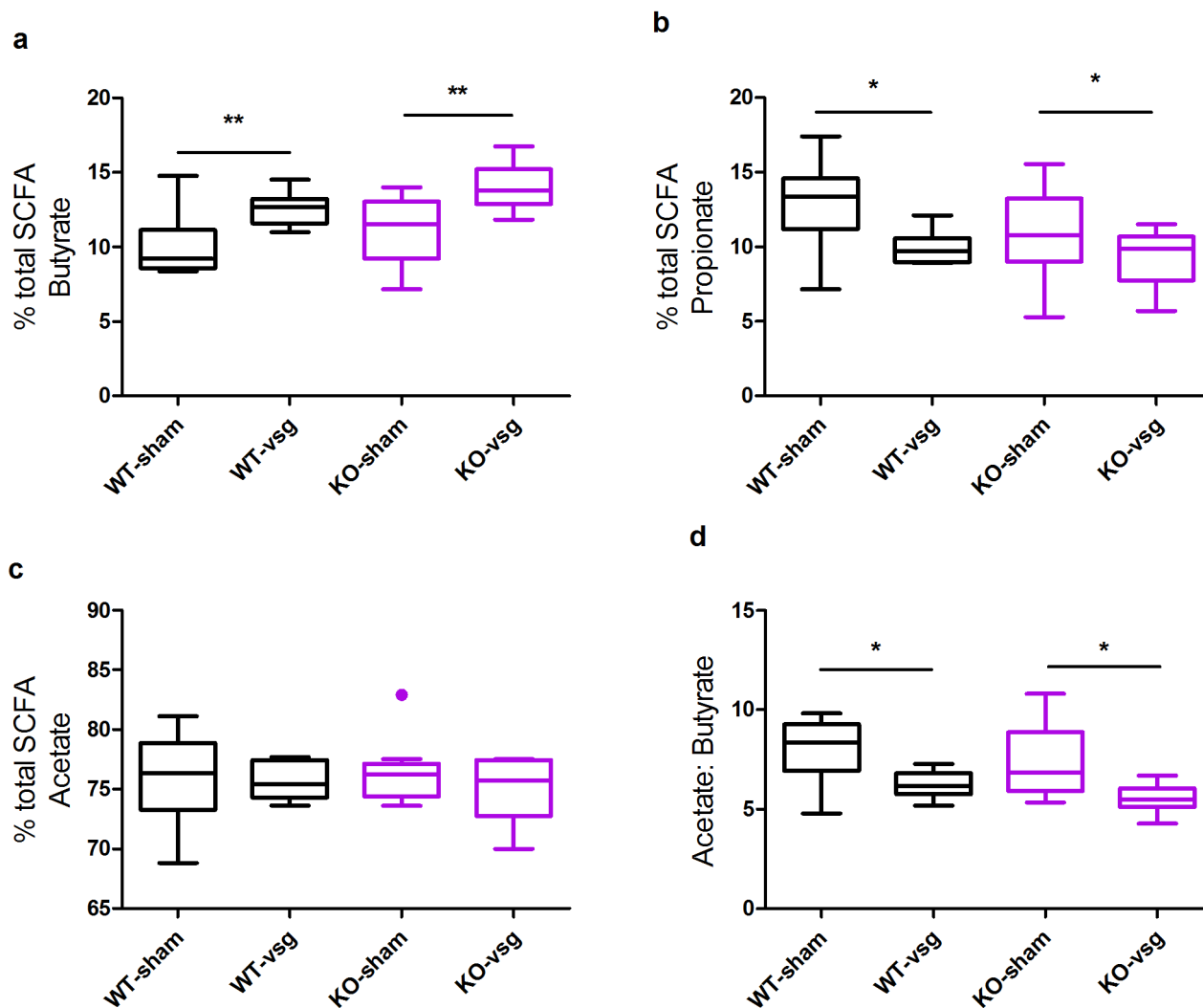
Extended Data Figure 4 | Relative abundance of *Bacteroides*, an uncharacterized genus in Porphyromonadaceae, and *Roseburia* correlated with metabolic parameters. a–d, The relative abundance of *Bacteroides* was significantly correlated with change in body weight (a), change in body fat (b), and the area under the curve (AUC) in the glucose tolerance test (d), but not with fasting blood glucose (c). e–h, The relative abundance of an

uncharacterized genus in Porphyromonadaceae was significantly correlated with change in body weight (e), fasting blood glucose (g) and AUC in the glucose tolerance test (h), but not with change in body fat (f). i–l, The relative abundance of *Roseburia* was significantly correlated with change in body weight (i), change in body fat (j), fasting blood glucose (k) and AUC in the glucose tolerance test (l). $n = 36$.



Extended Data Figure 5 | Effect of genotype and VSG on the relative abundance of *Lactobacillus*, *Lactococcus* and *Escherichia*. a–c, VSG was associated with a significant increase in the relative abundance of *Lactobacillus*

(a), *Lactococcus* (b) and *Escherichia/Shigella* (c) that did not vary according to genotype. Data are presented as Tukey box-plots. $n = 12$ WT-sham, 7 WT-VSG, 9 KO-sham, 8 KO-VSG.



Extended Data Figure 6 | VSG alters the abundance of caecal SCFAs.
a–c, The relative concentration of butyrate (**a**) and propionate (**b**), but not acetate (**c**), was altered by VSG, and this did not differ depending on genotype.

d, The acetate:butyrate ratio is increased following VSG. Data are presented as Tukey box-plots. * $P < 0.05$, ** $P < 0.01$. Also see Extended Data Table 1. $n = 12$ WT-sham, 7 WT-VSG, 9 KO-sham, 8 KO-VSG.

Extended Data Table 1 | The effect of VSG on caecal organic acids

Organic acid	WT-Sham	WT-VSG	KO-Sham	KO-VSG	p-value (2-way ANOVA)
Acetate ($\mu\text{mol/g}$)	142.33 \pm 20.45	153.46 \pm 26.77	124.54 \pm 23.61	160.86 \pm 25.04	ns
Propionate ($\mu\text{mol/g}$)	24.80 \pm 3.28	19.60 \pm 4.29	15.14 \pm 3.78	20.32 \pm 4.01	ns
Isobutyrate ($\mu\text{mol/g}$)	9.56 \pm 1.71	10.41 \pm 2.24	9.11 \pm 1.98	10.09 \pm 2.09	ns
Butyrate ($\mu\text{mol/g}$)	19.20 \pm 3.42	24.75 \pm 4.48	17.55 \pm 3.95	29.79 \pm 4.19	p (surgery) < 0.05
Isovalerate ($\mu\text{mol/g}$)	6.25 \pm 0.99	7.02 \pm 1.3	5.78 \pm 1.14	7.08 \pm 1.21	ns
Valerate ($\mu\text{mol/g}$)	2.50 \pm 0.41	2.77 \pm 0.54	2.19 \pm 0.41	2.99 \pm 0.41	ns
Hexanoate ($\mu\text{mol/g}$)	1.14 \pm 0.37	1.73 \pm 0.49	0.79 \pm 0.43	1.74 \pm 0.46	ns
Hepthanoate ($\mu\text{mol/g}$)	0.0080 \pm 0.0027	0.0088 \pm 0.0027	0.0032 \pm 0.0032	0.0088 \pm 0.0036	ns
Octanoate ($\mu\text{mol/g}$)	0.013 \pm 0.0056	0.027 \pm 0.0073	0.0073 \pm 0.0064	0.019 \pm 0.0068	ns
Lactate ($\mu\text{mol/g}$)	13.63 \pm 14.80	75.76 \pm 19.38	6.95 \pm 17.10	38.20 \pm 18.13	p (surgery) < 0.01
Succinate ($\mu\text{mol/g}$)	4.29 \pm 2.01	4.55 \pm 2.63	5.17 \pm 2.32	8.81 \pm 2.46	ns
Fumarate ($\mu\text{mol/g}$)	0.50 \pm 0.23	0.82 \pm 0.30	0.91 \pm 0.27	1.16 \pm 0.28	ns
Total SCFAs (C2-C4)	186.33 \pm 26.50	197.82 \pm 34.70	157.23 \pm 30.60	210.99 \pm 32.46	ns

Breakdown of the caecal organic acids ($\mu\text{mol per g dry weight}$) in both WT and KO mice after either sham or VSG surgery.

Astrocyte-encoded positional cues maintain sensorimotor circuit integrity

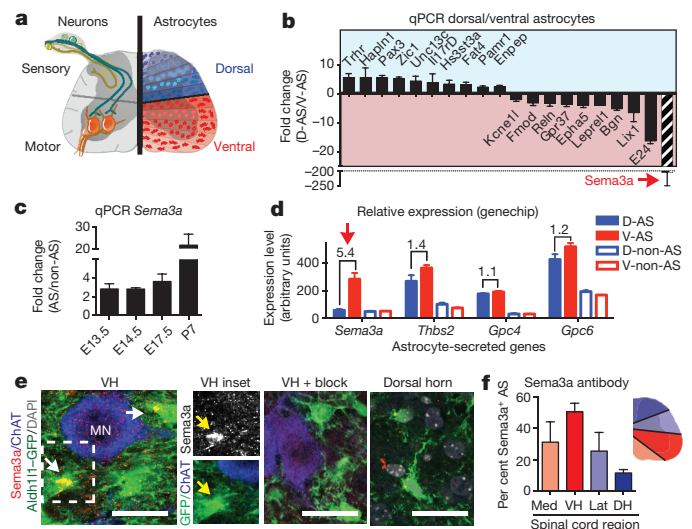
Anna V. Molofsky^{1,2,3}, Kevin W. Kelley^{1,2,4,5,6*}, Hui-Hsin Tsai^{1,2,4*}, Stephanie A. Redmond^{6,7}, Sandra M. Chang^{1,2}, Lohith Madireddy⁷, Jonah R. Chan⁷, Sergio E. Baranzini⁷, Erik M. Ullian⁸ & David H. Rowitch^{1,2,4,9}

Astrocytes, the most abundant cells in the central nervous system, promote synapse formation and help to refine neural connectivity. Although they are allocated to spatially distinct regional domains during development, it is unknown whether region-restricted astrocytes are functionally heterogeneous. Here we show that postnatal spinal cord astrocytes express several region-specific genes, and that ventral astrocyte-encoded semaphorin 3a (*Sema3a*) is required for proper motor neuron and sensory neuron circuit organization. Loss of astrocyte-encoded *Sema3a* leads to dysregulated α -motor neuron axon initial segment orientation, markedly abnormal synaptic inputs, and selective death of α - but not of adjacent γ -motor neurons. In addition, a subset of *TrkA*⁺ sensory afferents projects to ectopic ventral positions. These findings demonstrate that stable maintenance of a positional cue by developing astrocytes influences multiple aspects of sensorimotor circuit formation. More generally, they suggest that regional astrocyte heterogeneity may help to coordinate postnatal neural circuit refinement.

Developing neural circuits must form and maintain appropriate regional connections in a rapidly expanding central nervous system (CNS). Although astrocytes (AS) are increasingly recognized as general regulators of synapse formation¹, little is known about whether they encode heterogeneous positional signals involved in local neural circuit formation and/or maintenance. Recent studies indicate that AS develop and are regionally allocated in murine brain and spinal cord (SC) according to an embryonic segmental template^{2–4}. AS derived from embryonic radial glia⁵ migrate in the trajectory of these fibres and proliferate locally^{6,7}, yielding clonally related populations^{2,8} that retain spatial restriction into adulthood.

Here we tested whether AS allocated to discrete dorsal–ventral (DV) SC domains might be functionally adapted to support specific neural circuits and neuronal subtypes⁹. The SC sensorimotor circuit has well-defined organization in the DV axis (Fig. 1a). The ventral horn contains two types of motor neurons (MN), called α -MN and γ -MN, whose axons exit the ventral root to project to extrafusal (α) and intrafusal (γ) muscle fibres¹⁰. During development, afferent sensory fibres entering from the dorsal root ganglion (DRG) include type 1a proprioceptive afferents that synapse directly on ventral α -MN, and *TrkA*⁺ sensory axons that synapse in the dorsal grey matter¹¹.

Although programs that control MN diversification and connectivity are well established^{9,12}, comparatively little is known about non-neuronal signals that influence local circuit formation^{11,13}. We report that ventral AS-encoded semaphorin 3a (*Sema3a*), a secreted molecule that signals through plexin A/neuropilin 1 receptor (Nrp1) complexes^{14–16}, has critical roles in orienting MN axon initial segments (AIS), synapse regulation, MN subtype survival and normal patterning of a subset of *TrkA*⁺ sensory neurons. These findings establish a discrete molecular function for region-restricted AS.



AS express DV-restricted positional cues

To identify regionally distinct molecular differences we purified fibrous and protoplasmic AS from microdissected dorsal and ventral postnatal day 7 (P7) SC by flow cytometry using the *Aldh1l1-GFP* transgene reporter^{2,7,17,18} (Fig. 1a and Extended Data Fig. 1). Gene expression profiling and bioinformatic analysis identified 38 genes that were differentially expressed (Extended Data Fig. 1), and we validated these results by quantitative PCR. As shown (Fig. 1b), several genes encoded extracellular matrix molecules¹⁹ (*Hapln1*, *Pamr1*, *Enpep*, *Bgn*) or factors with positional roles in brain development, including reelin (*Reln*)²⁰ and Eph receptor A5 (*Epha5*; ref. 21). Of these, *Sema3a* was the most highly expressed ventral AS-specific gene, showing over threefold higher levels in radial glia and AS (versus non-AS) from embryonic day 13.5 (E13.5)–P7 (Fig. 1c), consistent with *Sema3a* expression *in situ* (Extended Data Fig. 2a–c)¹⁵. In contrast to *Sema3a*, genes for other AS-secreted molecules, including thrombospondin 2 (*Thbs2*) and glypican 4 and 6 (*Gpc4*, *Gpc6*), were expressed without positional distinction (Fig. 1d). *Sema3b*, *-c*, *-d*, *-e* and *-f* were low or undetectable in AS (data not shown).

AS *Sema3a* protein expression showed graded expression, with lowest numbers of *Sema3a*⁺ cells in the dorsal horn and highest numbers in the ventral horn. In ventral AS, *Sema3a* proteins appeared oriented towards MN soma (Fig. 1e, f). Although *Sema3a* messenger RNA transcripts were highly expressed in α -MN, we did not detect corresponding *Sema3a* protein levels (Fig. 1e and Extended Data Fig. 2e). MN express the obligate semaphorin receptor Nrp1 (refs 15, 22, 23) (Extended Data Fig. 2f). In addition, DRG TrkA⁺ sensory neurons, but not parvalbumin (PV)⁺ proprioceptive afferents, express high levels of Nrp1 (Extended

Data Fig. 2f–h). These findings suggested potential neuronal subtype-specific functions for AS-encoded *Sema3a*.

AS *Sema3a* restricts α -MN AIS orientation

We first investigated ventral astrocytic *Sema3a* function during early postnatal MN development. To conditionally target *Sema3a* in AS, we crossed *Sema3a*^{fl/fl} mice²⁴ to *hGFAPcre* mice²⁵, chosen because: (1) Cre activity commences in late-stage glial progenitors; (2) it targets protoplasmic AS; and (3) Cre activity segregates from SC MN²⁶ and interneurons (Extended Data Fig. 3a). *hGFAPcre:Sema3a*^{fl/fl} animals survived postnatally in near-normal numbers.

Previous work indicates that *Sema3a* has roles in supporting dendrite versus axon identity²⁷ and hippocampal neuron axon repulsion and dendrite growth *in vitro*²⁸. To assess a potential role for AS-encoded *Sema3a* in orienting MN axons *in vivo* we used ankyrin G (AnkG)²⁹ to define the direction of the AIS relative to the ventral root. By P7, large α -MN and smaller γ -MN normally exist in a 2:1 ratio in most MN pools¹⁰. To determine AIS orientation of both α -MN and γ -MN following loss of AS-encoded *Sema3a*, we used AnkG to identify the AIS and NeuN staining to distinguish α -MN (choline acetyltransferase (ChAT)⁺ NeuN⁺) from γ -MN (ChAT⁺ NeuN[−])¹⁰. This analysis was performed at both lumbar and cervical levels and the results were plotted on a positional grid (Fig. 2a–c and Extended Data Fig. 4). In control animals, we found that α -MN AIS were oriented an average of $39 \pm 33^\circ$ (s.d.) from the ventral root. By contrast, we observed marked disorganization of α -MN AIS orientation in *hGFAPcre:Sema3a*^{fl/fl} animals, with increases in both mean angle and variability ($62 \pm 46^\circ$ (s.d.); $P < 0.001$) (Fig. 2d and Extended Data Table 1) at all topographic

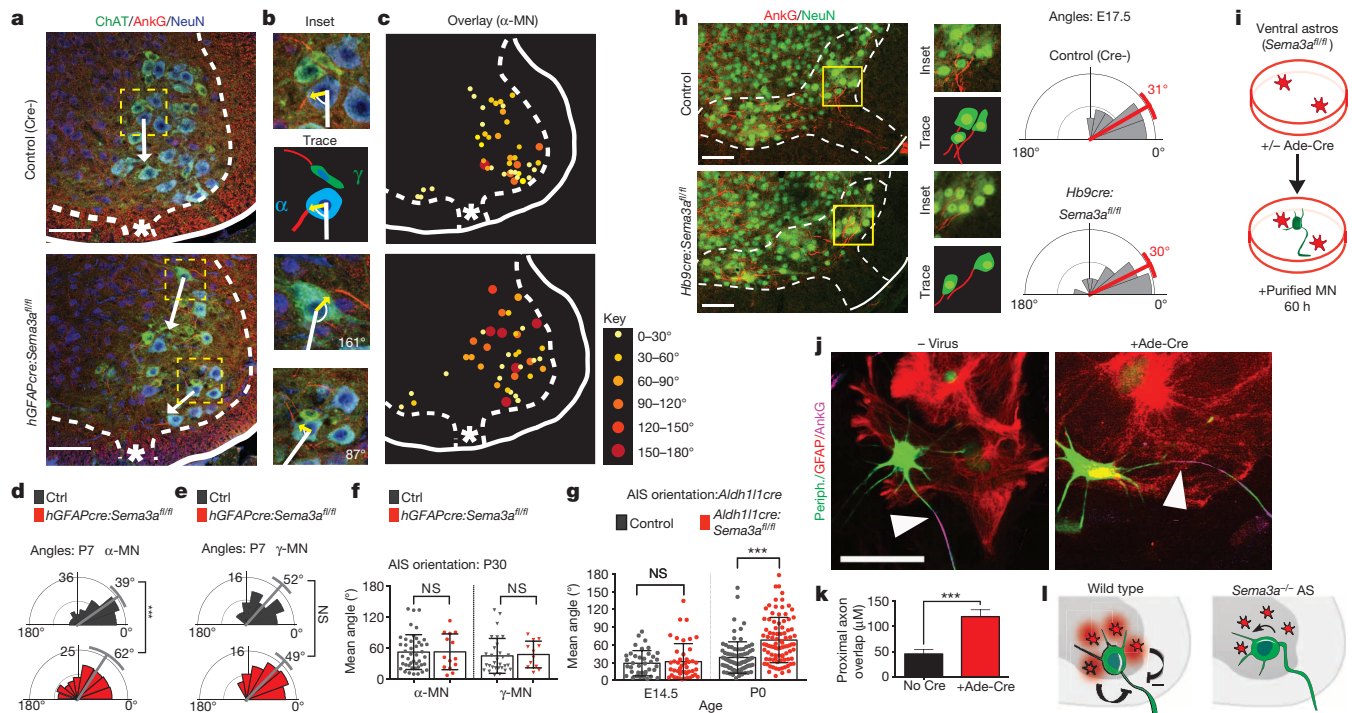


Figure 2 | Axon-repulsive effects of AS-encoded *Sema3a* maintain α -MN AIS orientation. **a**, P7 lumbar SC ChAT/NeuN⁺ α -MNs with AnkG⁺ AIS. Ventral root (*) and orientation vector towards it for selected α -MN (white arrows). Scale bar, 100 μ m. **b**, Insets of yellow boxed areas in **a**. Yellow arrow denotes AnkG⁺ AIS junction, the angle between these vectors determines AIS orientation. **c**, Overlay of all lumbar AIS orientation angles demonstrates that misoriented MN are topographically distributed. **d**, Mean angle and variability is significantly increased in P7 α -MN in absence of AS-encoded *Sema3a* (0° = towards ventral root). **e**, No difference in AIS orientation in γ -MN. **f**, Scatter plot of data generated as in **c** shows no misorientation of α - or γ -MN by P30. **g**, AIS angles in *Aldh1l1cre:Sema3a*^{fl/fl} are normal at E14.5

and significantly misoriented by P0. **h**, Deletion of *Sema3a* from MN with *Hb9cre* does not affect MN AIS orientation at E17.5. Scale bar, 50 μ m. **i**, AS-MN co-culture protocol. **j**, **k**, Increased MN axon overlap with Ade-Cre deletion of AS *Sema3a*^{fl/fl}. White arrowhead denotes proximal axon. Scale bar, 50 μ m. **l**, *Sema3a* secreted from ventral astrocytes maintains MN AIS orientation. Statistics and error bars: mean \pm s.d. Watson's U² test, except **k**, which shows mean \pm s.e.m, student's *t*-test. **d**, **e**, >100 neurons, $n = 3$ –4 per genotype, except at E14.5, which is $n = 2$ per genotype; **f**, >40 neurons, $n = 4$ per genotype; **h**, >30 neurons $n = 2$ per genotype; **k**, >20 neurons per condition, 3 independent experiments. *** $P < 0.001$; NS, not significant.

positions (Fig. 2c and Extended Data Fig. 4). Notably, we did not detect any differences in AIS orientation in γ -MN (control $52 \pm 31^\circ$ versus mutant $49 \pm 29^\circ$; $P =$ not significant; Fig. 2e). Furthermore, these prominent differences at P7 were no longer detectable by P30 (Fig. 2f).

To determine whether this AIS orientation defect reflected aberrant initial positioning or a failure to maintain MN orientation during later growth, we targeted radial glia and AS using *Aldh1l1cre*, which targets gliogenic radial glia at E12.5–13.5 (ref. 7). *Aldh1l1cre* fate maps most AS, some oligodendrocytes (that are *Sema3a* negative, Extended Data Fig. 2d), but excludes SC MN and interneurons (Extended Data Fig. 3b). *Aldh1l1cre:Sema3a^{fl/fl}* mice were perinatal lethal (28% expected survival at P1–P5, $n = 112$), but with normal numbers of AnkG⁺ MN, suggesting that axon specification is not affected *in vivo* (AnkG⁺ MN per confocal section: $42 \pm 5\%$ versus $46 \pm 4\%$, $P =$ not significant). At E14.5 (before the major period of AS expansion⁷), we found no evidence of abnormal MN cell body positioning (Extended Data Fig. 5) or MN AIS orientation in *Aldh1l1cre:Sema3a^{fl/fl}* animals (control $29 \pm 21^\circ$ s.d. versus mutant $32 \pm 30^\circ$, $P =$ not significant; Fig. 2g). By P0, AIS orientation defects were evident and similar in magnitude to those observed in P7 *hGFAPcre:Sema3a^{fl/fl}* ($37 \pm 25^\circ$ (s.d.) versus $66 \pm 38^\circ$ (s.d.); $P < 0.001$; Fig. 2g). As a control, we deleted *Sema3a* in MN using *Hb9cre^{30,31}*. This did not lead to defects in MN AIS orientation ($31 \pm 25^\circ$ control, $30 \pm 33^\circ$ *Hb9cre:Sema3a^{fl/fl}*, $P =$ not significant; Fig. 2h). These findings demonstrate that early developmental events—including initial positioning of MN soma and exit of MN axons from the ventral root—occur normally despite loss of AS-encoded *Sema3a*. However, with loss of AS-encoded *Sema3a*, many α -MN fail to properly maintain AIS orientation towards the ventral root at early postnatal stages.

To test whether AS-encoded *Sema3a* acts directly on MN without other cellular intermediates we co-cultured ventral SC AS from

Sema3a^{fl/fl} animals at sub-confluent density with embryonic rat MN³² (Fig. 2i), and added adenoviral Cre (Ade-Cre) to some wells to delete *Sema3a*. We found that the length of proximal segment overlap in MN adjacent to AS was significantly increased after Ade-Cre-mediated deletion of *Sema3a* from ventral AS ($50 \pm 9 \mu\text{m}$ versus $111 \pm 16 \mu\text{m}$, $P < 0.001$; Fig. 2j, k), suggesting a failure to properly repel the axon. Thus, AS-encoded *Sema3a* directly repels MN axons both *in vivo* and *in vitro*.

AS *Sema3a* is needed for α -MN survival

Given these findings, we investigated whether ventral AS-encoded *Sema3a* might also have later roles in maintenance of the sensorimotor circuit. At P7, numbers of both α - and γ -MN were normal in *hGFAPcre:Sema3a^{fl/fl}* mice (Fig. 3a, b¹⁰ and data not shown³³). By contrast, by P30 we found a significant ($P < 0.05$) twofold reduction in numbers of surviving Err3-negative¹⁰ α -MN in *hGFAPcre:Sema3a^{fl/fl}* mice (Fig. 3c, d). This was supported by a dose-dependent reduction in the average size of MN soma reflecting selective loss of a large MN population (Fig. 3e, f and Extended Data Table 1). By contrast, γ -MN persisted in normal numbers. In addition, two ventral interneuron populations (Chx10⁺ interneurons and calbindin⁺ Renshaw cells) did not show *Sema3a*-dependent depletion. In fact, the number of Renshaw cells was significantly increased at P30 (Extended Data Fig. 6).

To further assess direct effects of *Sema3a* on MN survival, we cultured MN in factor-free media for 24 h (to allow for initial polarization²⁸) then added recombinant *Sema3a* proteins. We found that exogenous *Sema3a* promoted MN survival in a dose-dependent manner that was abrogated by preincubating with an Nrp1-blocking antibody (Fig. 3g, h). Interestingly, the MN that survived in the absence of *Sema3a* typically had bipolar morphology (no add $86 \pm 7\%$ bipolar versus *Sema3a* $20 \pm 11\%$, $P < 0.01$). Together, these findings indicate that AS-encoded *Sema3a* can directly promote MN survival (Fig. 3i) in a manner that is tightly linked to its tropic effects. Notably, abnormal MN AIS orientation was no longer evident in adult *hGFAPcre:Sema3a^{fl/fl}* mice (Fig. 2f), suggesting that misoriented α -MN are lost by adulthood.

AS *Sema3a* regulates MN function

To test whether MN in *hGFAPcre:Sema3a^{fl/fl}* mutants integrated normally into local synaptic circuits (Fig. 4a), we first counted excitatory vGlut1⁺ presynaptic puncta on MN soma, which reflect type 1a proprioceptive sensory afferents¹¹. We observed a significant decrease in the number of vGlut1 puncta per MN (Fig. 4b; 4.82 ± 0.26 versus 2.91 ± 0.2 ; $P < 0.0001$) in adult mutant animals. These differences remained highly significant even when only large ($>500 \mu\text{m}$) putative α -MN were counted (5.59 ± 0.25 versus 4.13 ± 0.23 ; $P < 0.001$). Numbers of excitatory vGlut2 puncta, which are less dependent on DRG afferent input^{34–36}, were not significantly different in *hGFAPcre:Sema3a^{fl/fl}* animals versus controls (18.29 ± 0.93 versus 16.11 ± 0.88 , $P =$ not significant; Extended Data Table 1). By contrast, we found significant increases in numbers of VGAT inhibitory presynaptic puncta with loss of AS-encoded *Sema3a* (Fig. 4b; 21 ± 0.74 versus 27 ± 0.96 ; $P < 0.0001$).

To test whether these synaptic changes correlated with changes in MN function, we performed whole-cell patch clamp electrophysiological recordings³⁷ of lumbar α -MN at P13–14 from *hGFAPcre:Sema3a^{fl/fl}*; *Chat-GFP* transgenic reporter mice and Cre-negative controls (Fig. 4c, d). As shown (Fig. 4f–i), loss of AS-encoded *Sema3a* conferred a large shift in the balance of excitation/inhibition with significantly decreased MN spontaneous excitatory postsynaptic current (sEPSC) frequency (control 3.19 ± 0.35 Hz versus mutant 1.21 ± 0.38 Hz; $P < 0.01$) and increased spontaneous inhibitory postsynaptic current (sIPSC) frequency (control 0.14 ± 0.04 Hz versus mutant 0.84 ± 0.15 Hz; $P < 0.01$). Amplitudes of sEPSC and sIPSC were not changed (Extended Data Fig. 7). Interestingly, *hGFAPcre:Sema3a^{fl/fl}* MN were hyperexcitable as reflected by a significant decrease in the rheobase value (control 90.0 ± 23.0 pA versus mutant 37.5 ± 7.2 pA, $P < 0.05$; Fig. 4d, e) with unchanged input

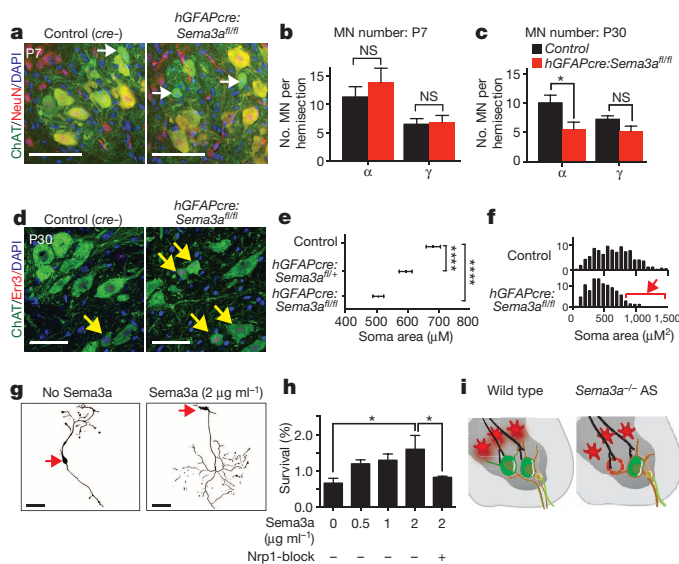


Figure 3 | AS-encoded *Sema3a* is required for postnatal α -MN survival. a, b, Normal numbers of α -MN (Chat⁺ NeuN⁺) and γ -MN (Chat⁺ NeuN⁻) at P7 in *hGFAPcre:Sema3a^{fl/fl}* cervical and lumbar SC. Scale bar, 50 μm . c, d, Fewer large α -MN and preserved numbers of γ -MN (Err3-bright, arrows.) in *hGFAPcre:Sema3a^{fl/fl}* animals at P28–P33. Scale bar, 50 μm . e, f, Dose-dependent decrease in MN soma area at P28–P33 with loss of AS-*Sema3a*; histograms show relative depletion of large MN (arrow). g, Representative peripherin⁺/Map2⁺ embryonic rat MN cultured in factor-free media or with recombinant *Sema3a*. Red arrow denotes soma. Scale bar, 100 μm . h, MN survival with recombinant *Sema3a* with/without Nrp1-blocking antibody. i, Summary. Statistics: mean \pm s.e.m. e, f, One-way mixed-effects analysis of variance (ANOVA) with Tukey's multiple comparison. b, c, h, Student's *t*-test. Data in b, c from 4 per genotype from 4 sections per animal; h average of 4 independent experiments. * $P < 0.05$, **** $P < 0.0001$.

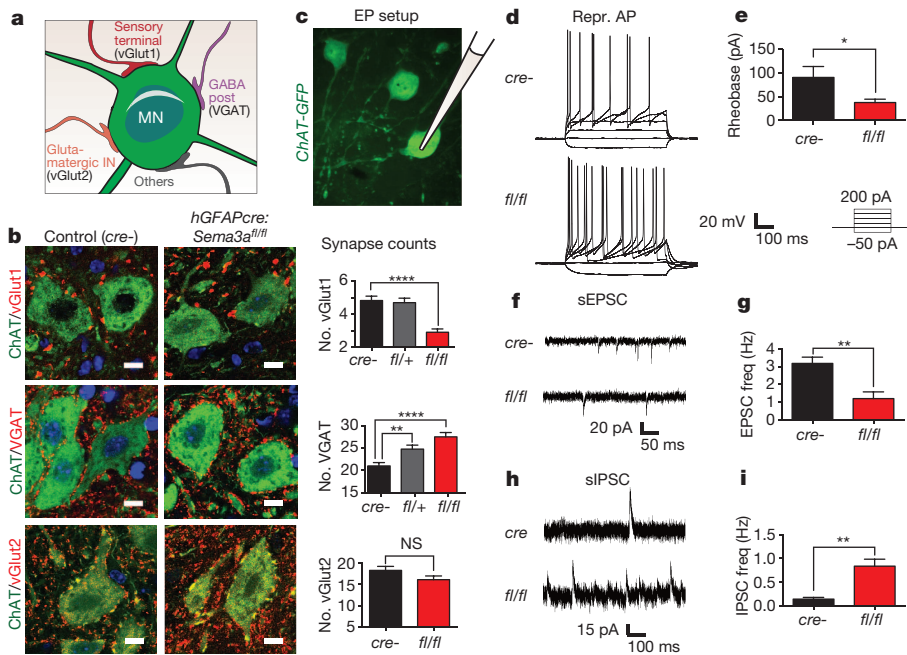


Figure 4 | AS-encoded *Sema3a* regulates MN synaptogenesis and function. **a**, Schematic of MN synaptic puncta. **b**, Decreased sensorimotor excitatory puncta (vGlut1⁺), increased inhibitory puncta (VGAT⁺), and preserved vGlut2⁺ inputs in hGFAPcre:Sema3a^{fl/fl} animals. Scale bar, 10 μm. **c**, Electrophysiology schematic. **d**, **e**, Representative action potentials and mean rheobase value demonstrate hyperexcitable hGFAPcre:Sema3a^{fl/fl} MN. **f**–**i**, Decreased sEPSC frequency (**f**, **g**) and increased sIPSC frequency (**h**, **i**) in hGFAPcre:Sema3a^{fl/fl} MN. Statistics: mean ± s.e.m. Data in **b** from cervical and lumbar levels of >4 per genotype and >200 MN per ea; data in **d**–**h** = 5–6 per genotype from lumbar slices. **b**, One-way mixed-effects ANOVA with Tukey's comparison; vGlut2: student's *t*-test. **d**–**h**, Student's *t*-test. **P* < 0.05, ***P* < 0.01, *****P* < 0.0001.

resistances (Extended Data Fig. 7), suggesting compensatory changes in MN function. Together, these findings demonstrate that loss of AS-encoded *Sema3a* leads to changes in MN excitatory and inhibitory synaptic inputs and has global effects on MN firing properties.

AS regulate sensory axon targeting

Initial studies of *Sema3a* demonstrated selective chemorepellent activity for sensory axons expressing TrkA or calcitonin gene-related peptide^{14,22,38}, which label overlapping sets of DRG neurons³⁹. However, the cellular source of *Sema3a* in SC has never been defined. Sensory axon guidance takes place between E13.5 and E18.5 in the mouse SC⁴⁰. As shown (Fig. 1a), TrkA⁺ sensory axons normally synapse in the dorsal horn⁴¹, whereas PV⁺ proprioceptive 1a afferents synapse with ventral α-MN¹¹.

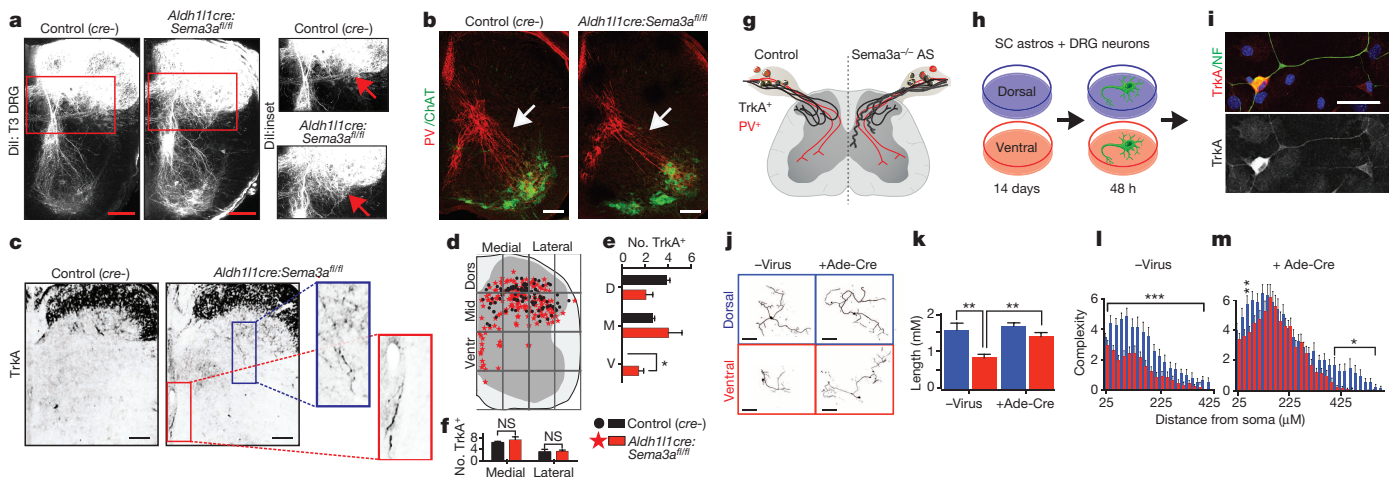


Figure 5 | AS-encoded *Sema3a* regulates DV positioning of sensory axons. **a**, DiI labelling of upper thoracic SC demonstrates ectopic ventral fibres in Aldh11cre:Sema3a^{fl/fl} mice (inset, box, arrows) (*n* = 7 per group). Scale bar, 100 μm. **b**, No ectopic proprioceptive 1a afferents (white arrow) in Aldh11cre:Sema3a^{fl/fl} mice; *n* = 4 per group. Scale bar, 100 μm. **c**, Multiple ectopic ventral TrkA⁺ projections in Aldh11cre:Sema3a^{fl/fl} mice (box, insets). Scale bar, 100 μm. **d**, Overlay dot-plot of TrkA⁺ terminations in Cre-negative controls (black circles) and Aldh11cre:Sema3a^{fl/fl} (red stars); *n* ≥ 3 per genotype. **e**, Quantification of DV termination index⁴⁴ shows a significant increase in ventral terminations in Aldh11cre:Sema3a^{fl/fl} mice. **f**, Medial versus lateral terminations unchanged. **g**, Summary of sensory axon phenotypes.

We characterized sensory axon guidance in Aldh11cre:Sema3a^{fl/fl} animals to determine potential AS-encoded functions (Fig. 5). Notably, whereas Aldh11cre activity fate maps to ~5% of DRG cells, *Sema3a* expression was undetectable in the DRG¹⁵ (Extended Data Fig. 2b), and we observed normal numbers of both PV⁺ and TrkA⁺ cells in the DRG of Aldh11cre:Sema3a^{fl/fl} mice (Extended Data Fig. 8). DiI labelling of DRGs of Aldh11cre:Sema3a^{fl/fl} mutants at E18.5–P0 showed abnormal ventrally positioned axons in 7 out of 7 mutants examined versus 0 out of 7 controls (Fig. 5a).

We then investigated the sub-classes of DRG sensory afferents affected in *Sema3a* mutants. We observed normal DV positioning of PV⁺ type 1a proprioceptive afferents albeit with subtle fasciculation defects (Fig. 5b). By contrast, TrkA⁺ afferents showed numerous abnormal

h, Diagram of AS-sensory neuron co-culture protocol. **i**, Representative TrkA⁺ DRG neuron grown on SC AS. Scale bar, 50 μm. **j**, Tracings of wild-type DRG neurons grown on dorsal/ventral SC AS from Sema3a^{fl/fl} mice, under wild-type (no virus) or Sema3a^{-/-} (+ Ade-Cre) conditions. Scale bar, 100 μm. **k**, Reduced total neurite length of neurons grown on ventral (red) vs dorsal (blue) AS significantly rescued on Sema3a^{-/-} ventral AS. **l**, **m**, Sholl analysis shows significantly less branching on ventral AS (I), rescued with Sema3a deletion (m). Statistics: culture data from 4–6 independent experiments with >10 neurons per condition per experiment. Mean ± s.e.m.; **P* < 0.05, ***P* < 0.01, ****P* < 0.001, Student's *t*-test.

ventral terminations in *Aldh1l1cre:Sema3a^{fl/fl}* mutants (Fig. 5c–f). Thus, AS-encoded *Sema3a* regulates DV patterning of sensory axon projections in a subtype-specific manner *in vivo* (Fig. 5g), consistent with previous results³⁸ and with the neuronal pattern of *Nrp1* expression (Extended Data Fig. 2).

To test whether distinctions between regional SC AS exist independent of environmental positional cues, we co-cultured dorsal or ventral AS with dissociated DRG sensory neurons (Fig. 5h) using AS from *Sema3a^{fl/fl}* animals and deleted *Sema3a* by adding *Ade-Cre*. Interestingly, AS cultures retained their distinct regional expression characteristics for many of the genes prospectively identified *in vivo* (see Fig. 1b and Extended Data Fig. 9), including higher ventral *Sema3a* levels.

We observed that both dorsal and ventral AS cultures preferentially supported the survival of TrkA-positive sensory neurons (>85% of neurons were TrkA⁺, none were PV⁺; Fig. 5i), with no significant survival differences between dorsal and ventral cultures. However, ventral SC AS cultures significantly ($P < 0.01$) inhibited neurite outgrowth and complexity relative to dorsal AS (Fig. 5j–m). *Ade-Cre* deletion of *Sema3a^{fl/fl}* normalized differences in neuronal length and complexity conferred by ventral AS (Fig. 5j–m). Together, these findings show that AS encode subtype-specific sensory axon guidance signals.

DISCUSSION

We propose that region-restricted AS comprise a stable ‘scaffold’ that maintains positional information throughout embryonic and postnatal development. This positional code is necessary for proper circuit formation, refinement and neuronal survival in a subtype-specific manner. Loss of *Sema3a* function from AS led to a sequence of α -MN-specific phenotypes, comprising defective postnatal maintenance of AIS orientation, markedly abnormal inhibitory and excitatory currents in MN and abnormal synapse investment, and finally, α -MN loss. Concomitantly, in more dorsal regions, AS-encoded *Sema3a* acts to repel TrkA sensory afferent fibres in a subtype-specific manner. It is possible that these phenotypes represent a pathological progression, or alternatively, that ventral AS-encoded *Sema3a* has multiple coordinated functions that determine structural and functional sensorimotor circuit integrity.

Our *in vitro* studies further suggest that AS positional identity is at least partly cell intrinsic, as cognate *Sema3a*-dependent regional AS properties were retained in co-cultures independent of local environmental cues. As such, a testable prediction is that embryonic CNS patterning mechanisms might establish a template for generation of heterogeneous properties of AS⁴². Furthermore, although maintenance of MN axonal orientation represents one tropic effect of AS-encoded *Sema3a*, further investigation is needed to assess other potential roles of this or other AS-encoded regional cues, such as promoting dendrite growth^{27,28}, maintenance of neuronal soma position, local synaptic strength and/or sensorimotor specificity¹³.

Specialized local functions of AS in neural circuit formation may also have significance in human disease. For example, loss of ventral SC MN in amyotrophic lateral sclerosis (ALS) has been associated with mutant superoxide dismutase protein in ventral AS in animal models of the disease⁴³. Our findings suggest the possibility that the unique identity of ventral horn AS might lead to deficient local support for MN and disease progression in ALS. More generally, given that AS are regionally patterned throughout the CNS², the concept of regional AS function and dysfunction has implications for a variety of neurodevelopmental and psychiatric disorders.

METHODS SUMMARY

Animals were maintained in the University of California San Francisco animal facility. All protocols were IRB approved and in accordance with the Institutional Animal Care and Use Committee guidelines. Circular data and statistics were analysed using Oriana4 software (Kovach Computing Services). Astrocyte monolayers from microdissected P0 mouse spinal cord were cultured for 12–14 days

before re-plating at sub-confluent density for MN co-cultures and high density for DRG co-cultures. Experiments with recombinant *Sema3a* (Peprotech) were on Matrigel in serum-free, growth-factor-free media. *Nrp1* blockade with 10 μ g ml⁻¹ antibody (R&D AF566) was added 30 min before *Sema3a*.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 8 August 2013; accepted 18 February 2014.

Published online 28 April 2014.

- Clarke, L. E. & Barres, B. A. Emerging roles of astrocytes in neural circuit development. *Nature Rev. Neurosci.* **14**, 311–321 (2013).
- Tsai, H.-H. *et al.* Regional astrocyte allocation regulates CNS synaptogenesis and repair. *Science* **337**, 358–362 (2012).
- Hochstim, C., Deneen, B., Lukaszewicz, A., Zhou, Q. & Anderson, D. J. Identification of positionally distinct astrocyte subtypes whose identities are specified by a homeodomain code. *Cell* **133**, 510–522 (2008).
- Muroyama, Y., Fujiwara, Y., Orkin, S. H. & Rowitch, D. H. Specification of astrocytes by bHLH protein SCL in a restricted region of the neural tube. *Nature* **438**, 360–363 (2005).
- Rowitch, D. H. & Kriegstein, A. R. Developmental genetics of vertebrate glial-cell specification. *Nature* **468**, 214–222 (2010).
- Ge, W.-P., Miyawaki, A., Gage, F. H., Jan, Y.-N. & Jan, L. Y. Local generation of glia is a major astrocyte source in postnatal cortex. *Nature* **484**, 376–380 (2012).
- Tien, A.-C. *et al.* Regulated temporal-spatial astrocyte precursor cell proliferation involves BRAF signalling in mammalian spinal cord. *Development* **139**, 2477–2487 (2012).
- Magavi, S., Friedmann, D., Banks, G., Stolfi, A. & Lois, C. Coincident generation of pyramidal neurons and protoplasmic astrocytes in neocortical columns. *J. Neurosci.* **32**, 4762–4772 (2012).
- Jessell, T. M. Neuronal specification in the spinal cord: inductive signals and transcriptional codes. *Nature Rev. Genet.* **1**, 20–29 (2000).
- Friesen, A. *et al.* Gamma and alpha motor neurons distinguished by expression of transcription factor *Err3*. *Proc. Natl Acad. Sci. USA* **106**, 13588–13593 (2009).
- Arber, S. Motor circuits in action: specification, connectivity, and function. *Neuron* **74**, 975–989 (2012).
- Dasen, J. S. & Jessell, T. M. Hox networks and the origins of motor neuron diversity. *Curr. Top. Dev. Biol.* **88**, 169–200 (2009).
- Sürmeli, G., Akay, T., Ippolito, G. C., Tucker, P. W. & Jessell, T. M. Patterns of spinal sensory-motor connectivity prescribed by a dorsoventral positional template. *Cell* **147**, 653–665 (2011).
- Messersmith, E. K. *et al.* Semaphorin III can function as a selective chemorepellent to pattern sensory projections in the spinal cord. *Neuron* **14**, 949–959 (1995).
- Kolodkin, A. L. *et al.* Neuropilin is a semaphorin III receptor. *Cell* **90**, 753–762 (1997).
- Pasterkamp, R. J. & Giger, R. J. Semaphorin function in neural plasticity and disease. *Curr. Opin. Neurobiol.* **19**, 263–274 (2009).
- Cahoy, J. D. *et al.* A transcriptome database for astrocytes, neurons, and oligodendrocytes: a new resource for understanding brain development and function. *J. Neurosci.* **28**, 264–278 (2008).
- Molofsky, A. V. *et al.* Expression profiling of *Aldh1l1*-precursors in the developing spinal cord reveals glial lineage-specific genes and direct Sox9-Nfe2l1 interactions. *Glia* **61**, 1518–1532 (2013).
- Barros, C. S., Franco, S. J. & Müller, U. Extracellular matrix: functions in the nervous system. *Cold Spring Harb. Perspect. Biol.* **3**, a005108 (2011).
- Tissir, F. & Goffinet, A. M. Reelin and brain development. *Nature Rev. Neurosci.* **4**, 496–505 (2003).
- Dufour, A. *et al.* Area specificity and topography of thalamocortical projections are controlled by ephrin/Eph genes. *Neuron* **39**, 453–465 (2003).
- Gu, C. *et al.* Neuropilin-1 conveys semaphorin and VEGF signaling during neural and cardiovascular development. *Dev. Cell* **5**, 45–57 (2003).
- Cohen, S. *et al.* A semaphorin code defines subpopulations of spinal motor neurons during mouse development. *Eur. J. Neurosci.* **21**, 1767–1776 (2005).
- Taniguchi, M. *et al.* Disruption of semaphorin III/D gene causes severe abnormality in peripheral nerve projection. *Neuron* **19**, 519–530 (1997).
- Zhuo, L. *et al.* hGFAP-cre transgenic mice for manipulation of glial and neuronal function *in vivo*. *Genesis* **31**, 85–94 (2001).
- McCall, M. A. *et al.* Targeted deletion in astrocyte intermediate filament (*Gfap*) alters neuronal physiology. *Proc. Natl Acad. Sci. USA* **93**, 6361–6366 (1996).
- Nishiyama, M. *et al.* Semaphorin 3A induces Ca_v2.3 channel-dependent conversion of axons to dendrites. *Nature Cell Biol.* **13**, 676–685 (2011).
- Shelly, M. *et al.* Semaphorin3A regulates neuronal polarization by suppressing axon formation and promoting dendrite growth. *Neuron* **71**, 433–446 (2011).
- Duflocq, A., Chareyre, F., Giovannini, M., Couraud, F. & Davenne, M. Characterization of the axon initial segment (AIS) of motor neurons and identification of a para-AIS and a juxtapara-AIS, organized by protein 4.1B. *BMC Biol.* **9**, 66 (2011).
- Arber, S. *et al.* Requirement for the homeobox gene *Hb9* in the consolidation of motor neuron identity. *Neuron* **23**, 659–674 (1999).
- Huber, A. B. *et al.* Distinct roles for secreted semaphorin signaling in spinal motor axon guidance. *Neuron* **48**, 949–964 (2005).

32. Camu, W. & Henderson, C. E. Purification of embryonic rat motoneurons by panning on a monoclonal antibody to the low-affinity NGF receptor. *J. Neurosci. Methods* **44**, 59–70 (1992).
33. Ashrafi, S. *et al.* *Wnt7A* identifies embryonic γ -motor neurons and reveals early postnatal dependence of γ -motor neurons on a muscle spindle-derived signal. *J. Neurosci.* **32**, 8725–8731 (2012).
34. Brumovsky, P., Watanabe, M. & Hökfelt, T. Expression of the vesicular glutamate transporters-1 and -2 in adult mouse dorsal root ganglia and spinal cord and their regulation by nerve injury. *Neuroscience* **147**, 469–490 (2007).
35. Brumovsky, P. R. VGLUTs in peripheral neurons and the spinal cord: time for a review. *ISRN Neurology* **2013**, 829753 (2013).
36. Alvarez, F. J., Villalba, R. M., Zerda, R. & Schneider, S. P. Vesicular glutamate transporters in the spinal cord, with special reference to sensory primary afferent synapses. *J. Comp. Neurol.* **472**, 257–280 (2004).
37. Mitra, P. & Brownstone, R. M. An *in vitro* spinal cord slice preparation for recording from lumbar motoneurons of the adult mouse. *J. Neurophysiol.* **107**, 728–741 (2012).
38. Behar, O., Golden, J. A., Mashimo, H., Schoen, F. J. & Fishman, M. C. Semaphorin III is needed for normal patterning and growth of nerves, bones and heart. *Nature* **383**, 525–528 (1996).
39. Averill, S., McMahon, S. B., Clary, D. O., Reichardt, L. F. & Priestley, J. V. Immunocytochemical localization of trkA receptors in chemically identified subgroups of adult rat sensory neurons. *Eur. J. Neurosci.* **7**, 1484–1494 (1995).
40. Ozaki, S. & Snider, W. D. Initial trajectories of sensory axons toward laminar targets in the developing mouse spinal cord. *J. Comp. Neurol.* **380**, 215–229 (1997).
41. Liu, Y. & Ma, Q. Generation of somatic sensory neuron diversity and implications on sensory coding. *Curr. Opin. Neurobiol.* **21**, 52–60 (2011).
42. Freeman, M. R. & Rowitch, D. H. Evolving concepts of gliogenesis: a look way back and ahead to the next 25 years. *Neuron* **80**, 613–623 (2013).
43. Nagai, M. *et al.* Astrocytes expressing ALS-linked mutated SOD1 release factors selectively toxic to motor neurons. *Nature Neurosci.* **10**, 615–622 (2007).
44. Chen, A. I., de Nooij, J. C. & Jessell, T. M. Graded activity of transcription factor Runx3 specifies the laminar termination pattern of sensory axons in the developing spinal cord. *Neuron* **49**, 395–408 (2006).

Acknowledgements We thank J. Flanagan, T. Jessell, J. de Nooij, N. Balaskas, M. Hancock, S. Ohata and R. Krencik for comments on the manuscript and technical suggestions. We are grateful to K. Sabeur, M. Wong and the UCSF Flow Cytometry and Genomics core facilities for expert technical help, A. Kolodkin for *Sema3a* probe construct, L. Reichardt for the TrkA antibody, J. Dasen for FoxP1 and Scip antibodies, and N. Heintz and J. Dougherty for *Aldh1L1*-cre mice. A.V.M. is supported by an NIMH Training Grant (5T32MH089920-04) and an APA/Pfizer MD/PhD Psychiatric Research Fellowship. K.W.K. is supported by the California Institute for Regenerative Medicine (TG2-01153). S.A.R. is supported by a Ruth L. Kirschstein NRSA FNS081905A. This work was supported by grants from the NINDS (to D.H.R. (R01 NS059893) and J.R.C.), E.M.U. is supported by NIMH (R01MH099595-01), an NIH New Innovator Award (1DP20D006507-01) and That Man May See. D.H.R. is a HHMI Investigator.

Author Contributions A.V.M. performed most experiments and data analysis. K.W.K. performed electrophysiology under supervision of E.M.U. H.-H.T. contributed to data analysis and experimental design. S.A.R. performed MN purification under supervision of J.R.C. S.M.C. performed mouse genotyping. L.M. and S.E.B. performed bioinformatics data processing and analysis. A.V.M. and D.H.R. designed the experiments and wrote the manuscript.

Author Information Microarray data has been deposited to GEO under accession number GSE55054. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to D.H.R. (rowitchd@peds.ucsf.edu).

METHODS

Mice. All mouse strains were maintained in the University of California San Francisco (UCSF) specific pathogen-free animal facility, and all animal protocols were approved by and in accordance with the guidelines established by the Institutional Animal Care and Use Committee and Laboratory Animal Resource Center. Mouse strains not otherwise referenced include CAG-GFP (MGI:3849685) and ChAT-eGFP (MGI:3694555). For embryonic tissues, plug date was considered as E0.5. Embryo age was confirmed by morphology and crown-rump length measurements at collection.

Astrocyte isolation by flow cytometry. Postnatal day 7 spinal cords were micro-dissected using an 'open book' preparation to separate dorsal and ventral halves. DRGs and meninges were removed then dissociated with papain (20 U ml⁻¹; Worthington) for 80 min at 33 °C as described previously¹⁷. Aldh1l1-positive and -negative cells were sorted as previously described¹⁸ on a BD FACS Aria II and gated on forward/side scatter, live/dead by DAPI exclusion, and GFP, using GFP-negative and DAPI-negative controls to set gates for each experiment. In some cases GFP-positive populations were re-sorted using the same gates to >95% purity.

RNA isolation. RNA was isolated using TRIZOL reagent (Invitrogen) with glycogen added as carrier, DNase-digested to remove genomic DNA contamination, and further purified using the RNeasy Kit (Qiagen). For microarray analysis, RNA samples were amplified using the Nugen Pico WT Ovation Kit and hybridized to Affymetrix Mouse Gene 1.0 ST arrays.

Bioinformatics. Microarray data were pre-processed in R using the Bioconductor suite of software packages. The 'oligo' package was used to background correct, normalize and summarize 1,102,500 probes on mouse gene 1.0 ST arrays via the robust multi-array analysis (RMA) algorithm. After nonspecific filtering was applied to remove low-intensity (<100 FU in at least 75% of the arrays), low-variance, and un-annotated probe sets, 7,799 probe sets remained. The Limma package was used to compare and assess differential expression between different groups of samples using the 'treat' algorithm. Dorsal astrocytes, ventral astrocytes, dorsal non-astrocytes, and ventral non-astrocytes were compared directly for the 'dorsal versus ventral' analysis. To generate an 'astrocyte versus non-astrocyte' data set, dorsal and ventral astrocyte probe sets were assigned equal weight and compared against dorsal and ventral non-astrocyte probe sets. With fold-change >1.2, false discovery rate = 0.15, we identified 5,158 genes differentially expressed between astrocytes and non-astrocytes, and 38 genes (39 transcripts) differentially expressed between dorsal and ventral astrocytes.

qPCR analysis. Complementary DNA was generated from purified RNA using Superscript III (Invitrogen) and random decamers. Primers were designed for amplicons of 75–150 bp using Primer 3 and are available upon request. qPCR was done on a Roche Lightcycler 480 using Sybr Green Master Mix (Roche). Melt curves were analysed for each experiment to ensure primer specificity. In most cases both β -actin (*Actb*) and *Gapdh* were used as housekeeping genes for normalization with similar results, and *Aldh1l1* was used as a housekeeping gene *in vitro*.

Immunohistochemistry/in situ hybridization. Most images were collected using a Leica SP5 confocal microscope. Antibodies used included mouse NeuN (Millipore), rabbit Sema3a (EMD biosciences), goat ChAT (Millipore), rabbit TrkA (gift of L. Reichardt, UCSF), chick GFP (Aveslabs), mouse Parvalbumin (Sigma), chick neurofilament (Encor), rabbit ankyrin G (Santa Cruz), mouse Err3 (PPMX), chick peripherin (Millipore), rabbit Map2 (Millipore) mouse Isl 1/2 (DSHB), sheep Chx10 (Abcam) and rabbit calbindin (Swant). Rabbit FoxP1 and guinea pig Scip were provided by J. Dasen. In most cases staining was done overnight at 4 °C in 5% serum/0.4% Triton, following heat-mediated antigen retrieval for 2 min at 95 °C in 0.1 M citrate buffer, pH 6.0. For GFP labelling antigen retrieval was 10 min at 70 °C. For TrkA staining, no retrieval was used, and slides were stained overnight at room temperature (20 °C) in 5% serum/1% Triton. *In situ* hybridization was performed using standard protocols. The probe for the full-length rat *Sema3a* transcript was provided by A. Kolodkin, the *Wnt7a* probe was from A. McMahon and *Plp1* probe was from I. Griffiths.

Polarity analysis in vivo. Measurements of motor neuron orientation relative to the ventral root were calculated by measuring the angle between the following: (1) the vector from the MN nucleus to the axon hillock, marked by AnkG staining of the proximal axon segment; and (2) the vector from the MN nucleus and the exit point of the ventral root from the grey matter. Angle measurements ranged from 0° (axon pointing towards the ventral root) to 180° (axon pointing directly away from ventral root). Circular data was analysed using Oriana4 software (Kovach Computing Services), and statistical analyses performed using a Watson's U² test.

Synapse and soma diameter counts. Counts of synaptic puncta on MN soma were performed at cervical (C4–5) and lumbar (L3–4) levels. Counts in the main figures represent pooled unbiased data from all levels with no size cutoff; however, all data were analysed in histogram format and sorted by MN soma area to determine whether results were likely to be biased by altered ratios of MN subtypes. For

some subgroup analyses (Extended Data Table 1), putative α -MN and γ -MN were identified by size on the basis of histogram analyses of soma area.

Astrocyte cell culture. Dorsal and ventral spinal cords from P0–1 mice were isolated and dissociated as above. Cells were plated at a density of $>1 \times 10^6$ per 25 cm² flask in DMEM-hi glucose with 10% FCS/10 μ M hydrocortisone, 5 μ g ml⁻¹ N-acetylcysteine, 2 μ g ml⁻¹ insulin and 20 ng ml⁻¹ EGF. Six days after plating cells, flasks were shaken to remove oligodendrocyte contamination. At 8 days, AraC was added to kill rapidly proliferating cells. For DRG co-cultures, 10–12 days after initial plating cells were re-plated into assay containers that consisted of 8-well glass chamber slides (BD) coated with poly-D-lysine and recombinant human fibronectin (Biomedical Technologies) to promote astrocyte adhesion, plating 30,000 cells per well. In most cases, cultures were established using *Sema3a*^{fl/fl} mice, and adenoviral Cre recombinase (Vector Biolabs) was added to some wells 2–4 h after re-plating. For MN co-cultures, astrocytes were re-plated at 2,000 cells per well onto a reduced GF Matrigel substrate (BD) diluted 1:25 in DMEM. Astrocyte monolayers were then cultured for 2–3 days before adding neurons.

DRG isolation and co-culture. DRGs from E13.5–14.5 mouse embryos were isolated and dissociated for 45 min in 0.25% trypsin/EDTA (Invitrogen). Five-hundred cells per well were plated onto astrocyte monolayers in minimal neural growth media containing DMEM:F12, 10% FCS, N2 and B27 supplements (Invitrogen), and co-cultured for 48 h before fixation in 4% paraformaldehyde and immunolabelling.

Motor neuron isolation and co-culture. Spinal cord neurons were isolated from embryonic rat spinal cords on the basis of previous protocols³². In brief, spinal cords were dissected from E15 rat embryos, dissociated in 0.25% trypsin (Gibco) for 15 min and triturated to form a single-cell suspension in L-15 plus 10% FBS media (Gibco). The suspension was immunopanned in a series of negative selection plates against rat neural antigen 2 (Ran2) and galactocerebroside, and then motor neurons were positively selected for on a final p75NTR panning plate. Adherent cells were released from the plate with a brief application of 0.05% trypsin (Gibco) and re-suspended in growth media (DMEM, B27, N2, Pen-Strep (Gibco)) before culturing. At re-plating (onto Matrigel-plated sub-confluent astrocytes for co-cultures, or Matrigel-coated wells for recombinant Sema3a experiments), fresh media was added consisting of DMEM hi-glucose supplemented with N2 and B27 supplements 5 μ g ml⁻¹ N-acetylcysteine, 5 μ g ml⁻¹ insulin and 5 μ M forskolin.

Recombinant Sema3a addition and Nrp1 blockade in vitro. MN were plated at 1,000 cells per well as above. 12 h after plating, they were assessed by light microscopy, at which point most could be seen to have budded polar processes. 24 h after plating, recombinant human Sema3a was added (Peprotech) at indicated concentrations. For Nrp1 blockade, an Nrp1-blocking antibody validated for this purpose (R&D Systems, AF566) was added at a concentration of 10 μ g ml⁻¹ and incubated with MN for 30 min before Sema3a addition. Goat IgG was added separately to control for nonspecific effects (not shown). Cells were cultured for an additional 36 h before fixation and immunolabelling.

Culture image analysis. For DRG co-cultures, neurons were identified with NF-H (Encor), in some cases TrkA, and PV immunostaining was used to identify sensory neuron subtypes. MN were identified with peripherin and traced with an overlay of peripherin and Map2 (to label distal dendrites). Spatially distinct neurons were photographed and analysed using NeuronJ plugin/Image J to trace total neurite length for each neuron. Sholl analyses were performed using the Sholl analysis plugin after thresholding each neuron using Image J (parameters: start 25 μ M, step size 10 μ M). More than ten neurons per condition per experiment were analysed. For axon overlap *in vitro*, only MN directly abutting an astrocyte were included in the analysis, and overlap of proximal axon (defined as the axon segment proximal to the first branch point) was quantified using ImageJ. For neurite length, statistical analyses were performed on means of each experiment, and in Sholl and axon overlap analyses, *t*-tests were performed on pooled data.

Whole-cell patch clamp recordings. Acute fresh lumbar (L3–4) spinal cord slices were prepared from *hGFAP-cre:Sema3a^{fl/fl}:ChAT-GFP* mice and Cre-negative controls from P12–14 using previously described protocols and solutions³⁷. In brief, transverse slices (350- μ m thick) were cut with a vibratome (Leica Microsystems) in a chamber filled with ice-cold sucrose cutting solution followed by a brief (60 s) incubation in polyethylene glycol (Mn = 1,900–2,200). The slices were then incubated in cutting solution at 35 °C for 30 min followed by 30 min in artificial cerebrospinal fluid then equilibrated to room temperature. Whole-cell recordings were made using patch clamp amplifiers (Multiclamp 700B) under an infrared-differential interference contrast microscope. Data acquisition and analysis were performed using digitizers (DigiData 1440A) and analysis software pClamp 10 (Molecular Devices). Signals were filtered at 6 kHz and sampled at 20 kHz. sEPSCs were recorded from motor neurons at –75 mV in voltage-clamp mode where the chloride reversal potential was 0 mV. sIPSCs were recorded at –55 mV where chloride currents were positive deflections in voltage clamp. Glass pipettes with a resistance of 2.5–4 M Ω were filled with a K-methanesulphonate internal solution. To ensure currents measured were sEPSCs and sIPSCs, control

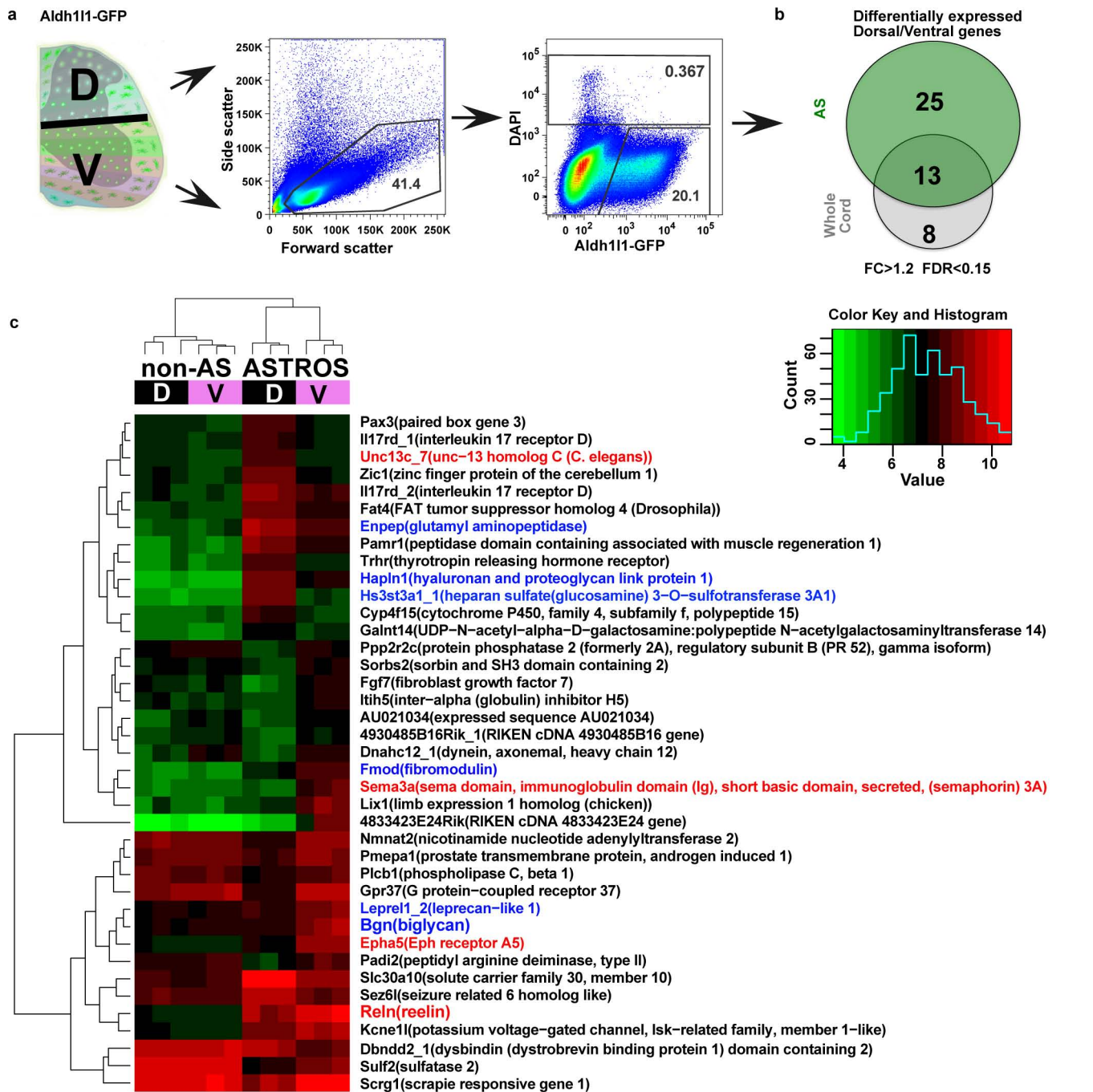
recordings were performed using standard pharmacology (Extended Data Fig. 7). Series resistance (15–25 M Ω) was monitored throughout the whole-cell recording and data were discarded if the change in series resistance was >20% during the course of the experiment.

DiI labelling of sensory afferents. Spinal cords from E18.5–P0 mice were fixed for at least 48 h in 4% paraformaldehyde, then dissected from spinal column with DRGs intact. The ventral root was severed and a crystal of DiI (Molecular probes) was applied to the DRG with a needle. Cords were incubated for 12 days at 37 °C, then 300- μ m thick sections were cut by vibratome, mounted in PBS, and imaged by collecting confocal z-stacks at 2 μ m spacing.

Sensory neuron termination analysis. Confocal sections of cervical spinal cords were overlaid with a uniform grid using Adobe Illustrator and terminations per

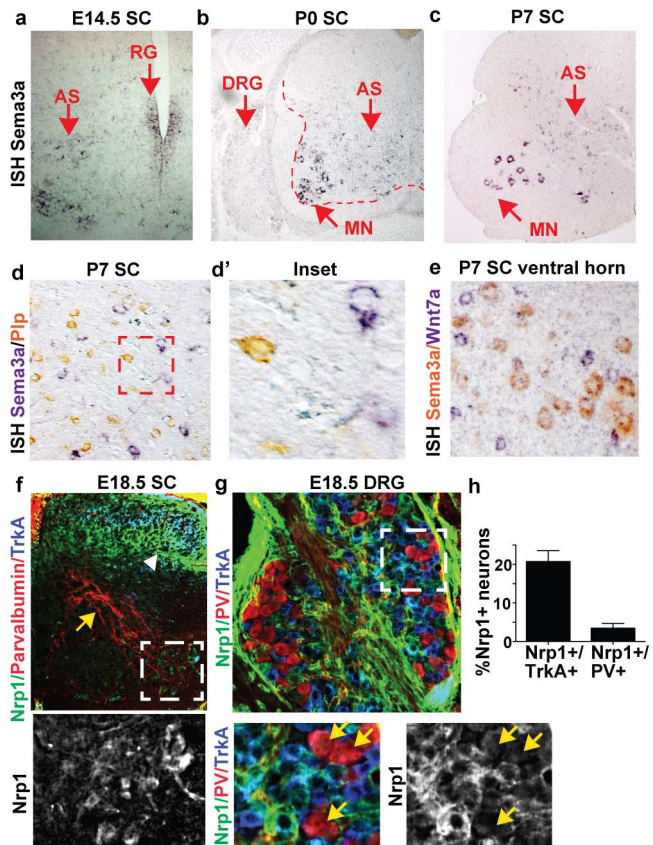
image were labelled. When more than one break in a process was visible, only the distal-most termination was labelled. Data from all sections and animals were overlaid on a normalized grid and the number of terminations per quadrant calculated.

Statistical analysis. Consultation obtained from the UCSF department of Biostatistics. Coefficient of variance (standard error as a percentage of the mean) was used to calculate minimum sample sizes. In most cases, sample sizes were well in excess. Student's *t*-tests were two-tailed and based on Gaussian distributions. In all cases, replicates refer to biological rather than technical replicates. Blinding during analysis was used whenever possible for all *in vivo* studies. Most statistics analysed using Graphpad Prism software. Circular data in Fig. 2 were analysed using Oriana4 software (Kovach Computing Services).

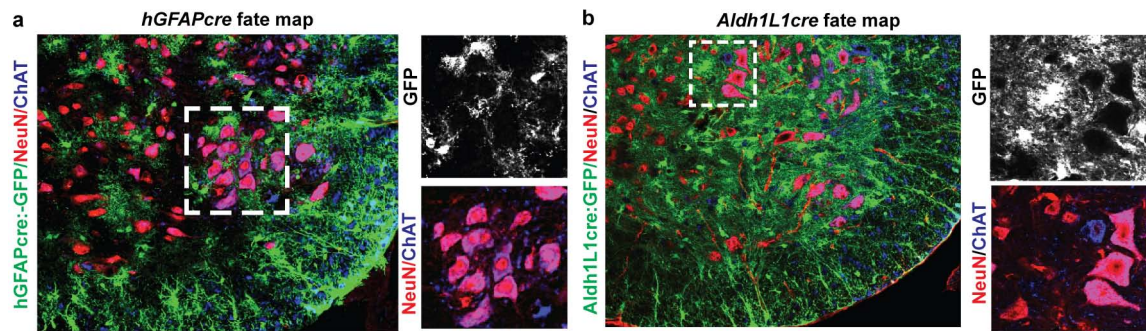


Extended Data Figure 1 | Flow cytometry gating strategy and microarray.
a, Schematic indicating microdissection of Aldh1l1-GFP-positive P7 spinal cord and isolation by flow cytometry using scatter gates, doublet exclusion (not shown) and sorting for GFP-positive cells with live/dead exclusion by DAPI staining. Percentage of Aldh1l1-GFP cells was not significantly different

between dorsal and ventral (not shown). b, Summary of differentially expressed genes in astrocytes (AS), whole cord, or both using the analysis parameters indicated. c, Heatmap of all 39 genes differentially expressed between dorsal and ventral cord, highlighting astrocyte-enriched genes with known roles in neural circuit development (red) or extracellular matrix (blue).

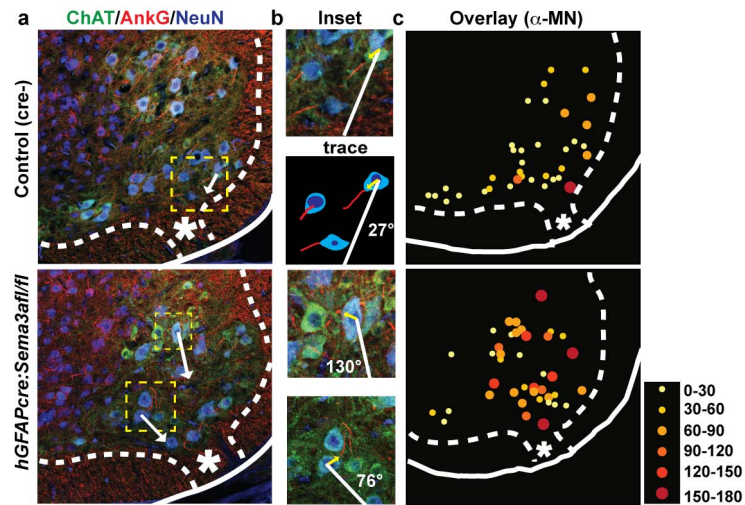


Extended Data Figure 2 | Coordinate expression of Sema3a and Nrp1 in astrocytes and neurons. **a–c**, *Sema3a* mRNA is expressed in radial glia (RG) and in protoplasmic cells that are NeuN negative throughout the embryonic and early postnatal period. *Sema3a* was not detected in DRG or in SC white matter (**b**). **d**, *Sema3a* is segregated from *Plp*-positive oligodendrocytes. **e**, MN *Sema3a* expression is detected in α -MN but not γ -MN in cervical SC. **f, g**, High levels of Nrp1 expression in TrkA⁺ fibres and cell bodies (white arrowhead) and in MN, but not in PV-positive fibres and cell bodies (yellow arrows). **h**, Quantification of percentage of Nrp1⁺ neurons per condition.



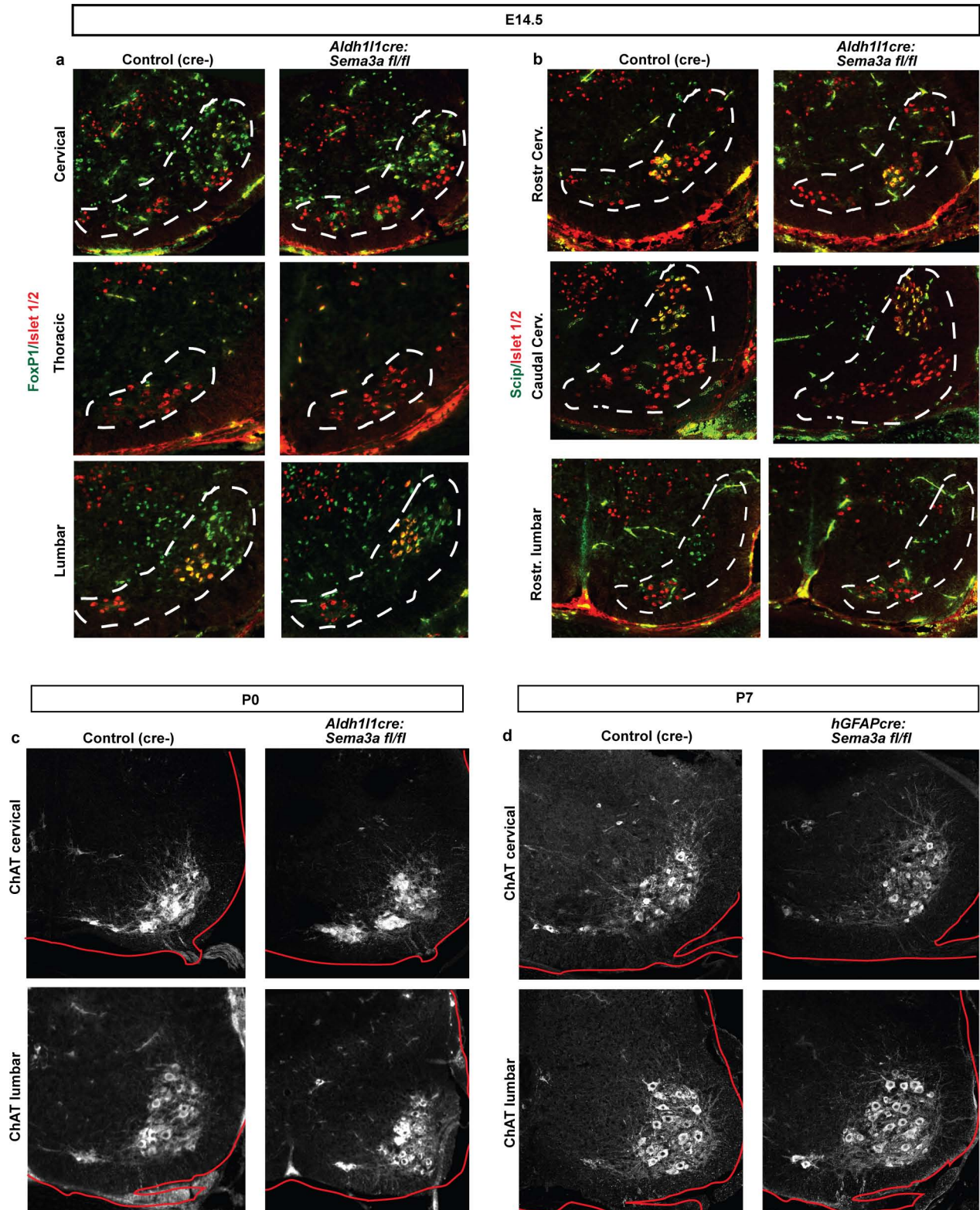
Extended Data Figure 3 | Fate map of conditional astrocyte deletion lines used in this study. **a**, *hGFAPcre* fate map labels fibrous and a subset of protoplasmic AS but not MN or interneurons in P10 SC. **b**, *Aldh1L1cre* fate

maps to astrocytes but not to neurons in P10 SC, including α -MN (purple), γ -MN (blue) and interneurons (red).



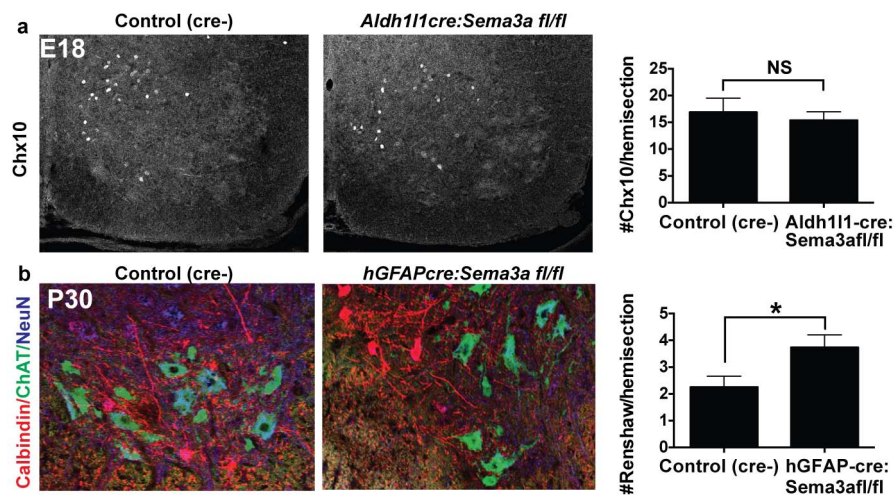
Extended Data Figure 4 | Motor neuron AIS orientation defects in cervical spinal cord. **a**, Representative images of cervical SC confocal sections stained to distinguish α - and γ -MN and identify their proximal axon segment (asterisk denotes ventral root). **b**, Inset shows high-magnification view of

representative MN with identifiable AIS and a schematic of their location with respect to the ventral root. **c**, Overlay of all cervical α -MN angles measured to generate data summarized in Fig. 2c, with positional information preserved, demonstrates that misoriented AIS can be seen at all DV positions.



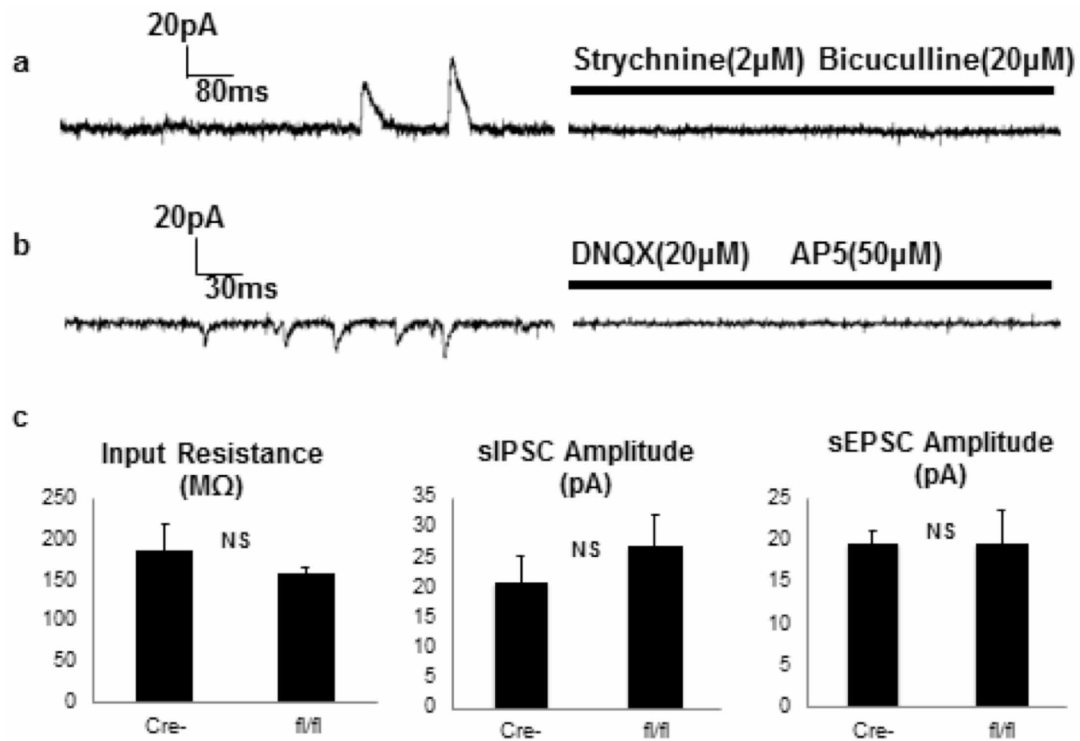
Extended Data Figure 5 | No evidence of abnormal MN cell body positioning with loss of astrocyte-encoded Sema3a. **a**, Representative FoxP1/Islet1/2 co-labelling at three rostrocaudal levels in control and mutant animals shows no differences between control and mutant. **b**, Similar stainings

using Scip (a PMC and LMC marker). **c**, **d**, No obvious differences in DV or mediolateral boundaries of ChAT⁺ MN at comparable cervical or lumbar levels at P0 (using *Aldh111cre* to delete Sema3a) and P7 (with *hGFAPcre*), both time periods where misorientation of AIS is clearly evident.



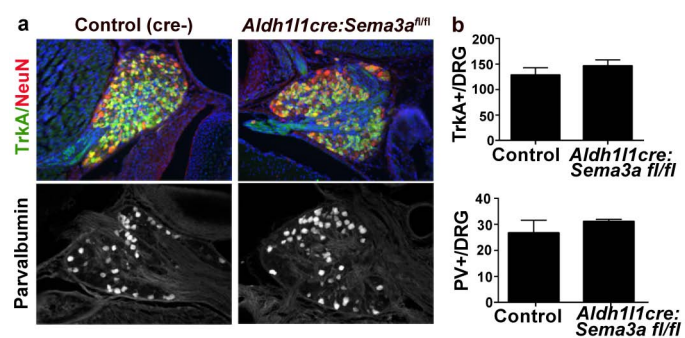
Extended Data Figure 6 | Quantification of ventral interneuron populations after loss of astrocyte-encoded Sema3a. **a**, Chx10 staining at E18 and quantification. **b**, Calbindin staining of Renshaw interneurons at P30 and quantification.

quantification demonstrates a significant increase at this age. Data are mean \pm s.e.m., student's *t*-test. Data in **a** from $n = 2$ per group, 4 sections per animal; data in **b** from 4 per group, 4 sections per animal.

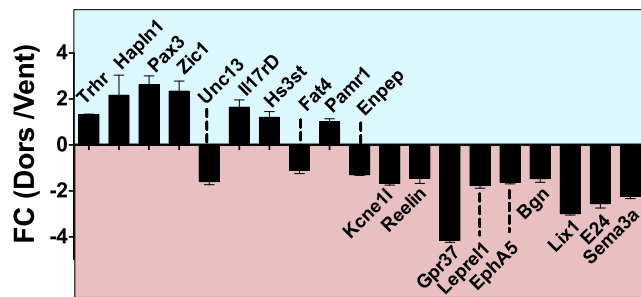


Extended Data Figure 7 | Additional data and controls for MN electrophysiology. **a**, 2 μM strychnine and 20 μM bicuculline block postsynaptic currents (at -55 mV) in a ChAT-GFP⁺ lumbar MN. **b**, 20 μM 6,7-dinitroquinoxaline-2,3-dione (DNQX) and 50 μM

(2R)-amino-5-phosphonovaleric acid (AP5) block postsynaptic currents (at -75 mV) in a ChAT-GFP⁺ lumbar MN. **c**, No difference in input resistance, sIPSC amplitude or sEPSC amplitude between control (Cre-) and *hGFAPcre:Sema3a^{fl/fl}* (fl/fl) MN. *n* = 5/each; mean ± s.e.m., Student's *t*-test.



Extended Data Figure 8 | Normal dorsal root ganglia in *Aldh1l1cre:Sema3a^{fl/fl}* mice. **a**, No difference in the number of subtype-specific neurons per DRG in control or *Aldh1l1cre:Sema3a^{fl/fl}* mice ($n = 3$ from 4–5 sections per animal; mean \pm s.e.m.; Student's t -test).



Extended Data Figure 9 | Differential expression of regionally heterogeneous astrocyte genes is partly preserved *in vitro*. qPCR quantification demonstrates that many regionally heterogeneous microarray genes prospectively identified *in vivo* remain differentially expressed *in vitro* after 17 days in culture, including ventral *Sema3a*. Mean \pm s.e.m., $n = 3$ independent experiments.

Extended Data Table 1 | Subgroup analyses of motor neuron data presented in Figs 2 and 3.

		Control (cre-)	<i>hGFAPcre: Sema3a^{fl/fl}</i>	<i>p</i> -value
MN axon orientation, P7 (° from ventral root)	All MN	42±33°(SD)	58±42°(SD)	**
	Cervical MN	42±34°(SD)	62±49°(SD)	**
	Lumbar MN	42±32°(SD)	55±42°(SD)	*
	α-MN (NeuN+)	39±33°(SD)	62±46°(SD)	***
	γ-MN (NeuN-)	52±31°(SD)	49±29°(SD)	ns
MN soma size, P30 (μm ²)	All MN	664±14	483±10	****
	Cervical MN	599±24	497±16	***
	Lumbar MN	687±17	469±14	****
#vGlut1/MN	All MN	4.8±0.26	2.9±0.20	****
	Cervical MN	4.3±0.29	2.9±0.19	***
	Lumbar MN	5.1±0.33	3.3±0.33	***
	α-MN (>500 μM)	5.6±0.25	4.1±0.24	***
	γ-MN (<500 μM)	2.27±0.28	1.95±0.19	ns
#VGAT/MN	All MN	21±0.74	27±0.96	****
	Cervical MN	19.88±1.1	24.5±1.5	*
	Lumbar MN	21.8±1.0	27.8±1.4	**
	α-MN (>500 μM)	25.7±1.1	37.5±1.3	****
	γ-MN (<500 μM)	11.4±0.82	22.7±1.0	***
#vGlut2/MN	All MN	18.29±0.93	16.11±0.88	ns
	Cervical MN	15.0±0.82	13.0±0.68	ns
	Lumbar MN	23.5±1.4	21.7±1.4	ns
	α-MN (>500 μM)	22.98±1.0	21.21±1.2	ns
	γ-MN (<500 μM)	12.18±0.81	11.7±0.57	ns

Synapse elimination and learning rules co-regulated by MHC class I H2-D^b

Hanmi Lee¹, Barbara K. Brott¹, Lowry A. Kirkby², Jaimie D. Adelson¹, Sarah Cheng¹, Marla B. Feller², Akash Datwani^{1†} & Carla J. Shatz¹

The formation of precise connections between retina and lateral geniculate nucleus (LGN) involves the activity-dependent elimination of some synapses, with strengthening and retention of others. Here we show that the major histocompatibility complex (MHC) class I molecule H2-D^b is necessary and sufficient for synapse elimination in the retinogeniculate system. In mice lacking both H2-K^b and H2-D^b ($K^bD^{b-/-}$), despite intact retinal activity and basal synaptic transmission, the developmentally regulated decrease in functional convergence of retinal ganglion cell synaptic inputs to LGN neurons fails and eye-specific layers do not form. Neuronal expression of just H2-D^b in $K^bD^{b-/-}$ mice rescues both synapse elimination and eye-specific segregation despite a compromised immune system. When patterns of stimulation mimicking endogenous retinal waves are used to probe synaptic learning rules at retinogeniculate synapses, long-term potentiation (LTP) is intact but long-term depression (LTD) is impaired in $K^bD^{b-/-}$ mice. This change is due to an increase in Ca²⁺-permeable AMPA (α -amino-3-hydroxy-5-methyl-4-isoxazole propionic acid) receptors. Restoring H2-D^b to $K^bD^{b-/-}$ neurons renders AMPA receptors Ca²⁺ impermeable and rescues LTD. These observations reveal an MHC-class-I-mediated link between developmental synapse pruning and balanced synaptic learning rules enabling both LTD and LTP, and demonstrate a direct requirement for H2-D^b in functional and structural synapse pruning in CNS neurons.

Early in development, before photoreceptors function, retinal ganglion cells (RGCs) spontaneously generate correlated bursts of action potentials called 'retinal waves'^{1–3}. Postsynaptic LGN neurons in turn are driven to fire in similar patterns^{4,5}, and this endogenous activity is even relayed further into the visual cortex⁶. Although there is consensus that retinal waves and correlated activity are needed for RGC synapse remodelling and segregation of RGC axons into eye-specific layers^{7,8}, little is known at the synaptic or molecular level about how natural patterns of activity are read out to drive elimination and structural remodelling before sensory experience. It is assumed that synaptic learning rules are present at retinogeniculate synapses, and that implementation of these rules ultimately leads either to synapse stabilization or elimination. Efforts to discover molecular mechanisms of developmental synapse elimination have implicated several unexpected candidates, all with links to the immune system, including neuronal pentraxins, Complement C1q and MHC class I family members^{9–11}. However, it is not known whether any of these molecules regulate plasticity rules at developing synapses. Moreover, because germline knockout mice were examined in each of these examples, it is not known whether neuronal versus immune function is required for synapse elimination *in vivo*. Here we examine these questions and also test whether genetically restoring H2-D^b expression selectively to CNS neurons *in vivo* can rescue synapse elimination in mice that nevertheless lack an intact immune system.

Defective synapse elimination in LGN of $K^bD^{b-/-}$ mutant mice

MHC class I genes *H2-D^b* and *H2-K^b*, members of a polymorphic family of over 50, are expressed in LGN neurons¹⁰ and were discovered in an unbiased screen *in vivo* for genes regulated by retinal waves: blocking this endogenous neural activity not only prevents RGC axonal remodelling⁷, but also downregulates expression of MHC class I messenger RNA¹². Previous studies have suggested that MHC class I molecules regulate synapse

number in cultured neurons¹³ and are needed for anatomical segregation of RGC axons into LGN layers *in vivo*^{10,14}. To examine if H2-K^b and H2-D^b are involved in functional synapse elimination, whole-cell microelectrode recordings were made from individual neurons in wild-type or $K^bD^{b-/-}$ LGN slices (Fig. 1a)^{15,16}. Adult mouse LGN neurons normally receive strong monosynaptic inputs from 1–3 RGC axons, but in development, many weak synaptic inputs are present. Most are eliminated between postnatal day 5 (P5) and P12 before eye opening, while the few remaining inputs strengthen, resulting in adult-like synaptic innervation by P24–P30 (ref. 15). By gradually increasing optic tract stimulation intensity, individual RGC axons with progressively higher firing thresholds can be recruited¹⁵, generating a stepwise series of excitatory postsynaptic currents (EPSCs) recorded in each LGN neuron. For example, at P21 in wild type, only two steps are present (Fig. 1b), indicating that just two RGC axons provide input to this LGN neuron, as expected. In contrast, in $K^bD^{b-/-}$ LGN neurons, there are many EPSC steps (Fig. 1c), a pattern similar to that in much younger wild-type mice before synapse elimination^{15,16}.

To obtain more quantitative information, minimal stimulation was used to estimate single fibre strength (SF-AMPA)¹⁷ (Methods and Extended Data Fig. 1a, b). On average, the amplitude of SF-AMPA in $K^bD^{b-/-}$ neurons is almost half that of wild type, and the cumulative probability distribution of EPSC amplitudes recorded from $K^bD^{b-/-}$ LGN neurons is also consistent with the presence of smaller sized EPSCs (Fig. 1d; note onset latency of SF-AMPA is similar in both genotypes (Extended Data Fig. 1c)). In contrast, maximal synaptic input (Max-AMPA) is not different between wild type and $K^bD^{b-/-}$ (Extended Data Fig. 1d). Fibre fraction, an index of how much each input contributes to total synaptic response¹⁵ (Methods), is half as large in $K^bD^{b-/-}$ than wild type (Fig. 1e), consistent with the idea that the number of RGC synapses in $K^bD^{b-/-}$ LGN neurons is greater than in wild type. An alternative possibility—that differences can arise from altered probability of

¹Departments of Biology and Neurobiology and Bio-X, James H. Clark Center, 318 Campus Drive, Stanford, California 94305, USA. ²Department of Molecular and Cell Biology & Helen Wills Neuroscience Institute, University of California, Berkeley, California 94720, USA. †Present address: Sage Bionetworks, 1100 Fairview Avenue N., Seattle, Washington 98109, USA.

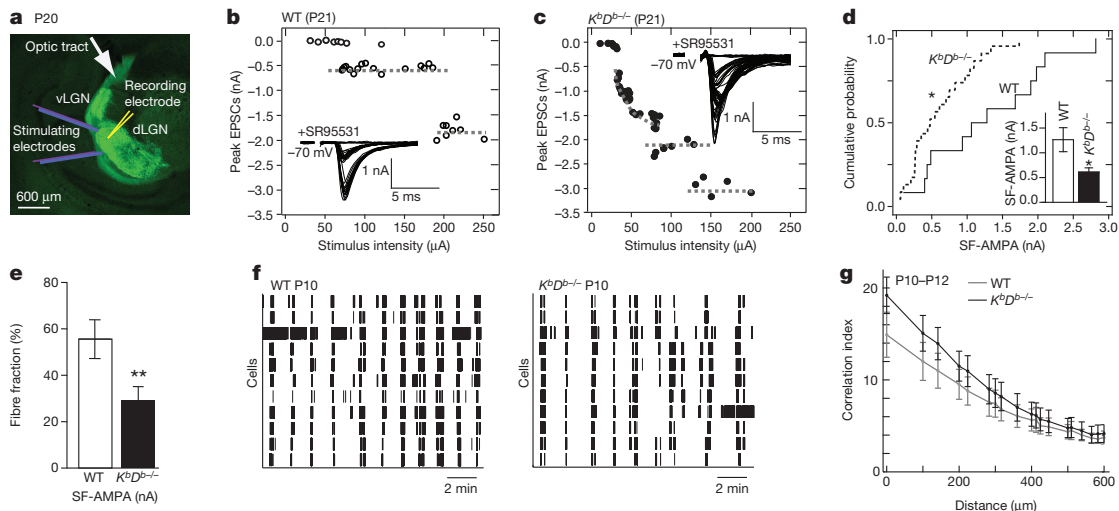


Figure 1 | Failure of retinogeniculate synapse elimination despite intact retinal waves in $K^bD^b-/-$. **a–e**, Impaired synapse elimination in $K^bD^b-/-$ mice at P20–P24. **a**, Slice preparation used for whole-cell recording from dLGN neurons and stimulation of retinal ganglion cell (RGC) axons in the optic tract. The retinogeniculate projection is visualized by injecting CTB AF488 (green) into the contralateral eye. **b**, **c**, EPSC amplitude versus optic tract stimulus intensity. Insets: example traces. **d**, Cumulative probability histograms of single fibre synaptic strength (SF-AMPA). Inset: mean \pm s.e.m. for wild type (WT) ($n = 12/N = 6$); $K^bD^b-/-$ ($n = 23/N = 8$), $*P < 0.05$. **e**, Fibre fraction for wild type ($n = 12/N = 6$); $K^bD^b-/-$ ($n = 21/N = 8$), $**P < 0.01$, t -test for **d**, **e**, **f**, **g**, Intact retinal waves in $K^bD^b-/-$ at P10–P12. **f**, Raster plots of single-unit spike trains recorded from 10 representative RGCs during retinal waves. **g**, Correlation indices versus inter-electrode distance for all cell pairs for wild type ($N = 5$) versus $K^bD^b-/-$ ($N = 6$). Data correspond to mean values of medians from individual data sets and error bars represent s.e.m. $n = \text{cells}/N = \text{animals}$.

release—is unlikely because paired-pulse ratio, an index of presynaptic release probability, is similar in wild type and $K^bD^b-/-$ at a variety of stimulus intervals (Extended Data Fig. 1e–h). Together, these experiments, which directly measure the functional status of synaptic innervation, demonstrate that either or both H2-K^b and H2-D^b are required for retinogeniculate synapse elimination.

Intact retinal wave activity in $K^bD^b-/-$ mice

Many previous studies have shown that retinogeniculate synapse elimination and eye-specific segregation in LGN fail if retinal waves are blocked or perturbed^{2,7,8}. Thus, waves could be absent or abnormal in $K^bD^b-/-$ mice. To examine this possibility, waves were recorded using a multi-electrode array to monitor action potential activity from many ganglion cells in $K^bD^b-/-$ or wild-type retinas between P5–P12, the peak period of extensive RGC synapse remodelling requiring waves. The spatio-temporal pattern of waves in $K^bD^b-/-$ is indistinguishable from wild type (Fig. 1f and Extended Data Fig. 2a–e). Moreover, the correlation index between all RGC pairs, a measure of the distance over which cells fire together^{1,3,18}, is almost identical (Fig. 1g and Extended Data Fig. 2a). Retinal wave activity also transitioned normally from cholinergic-dependent stage II (P5–P8) to glutamatergic-dependent stage III (P10–P12) (Extended Data Fig. 2a–e)⁸. After eye opening, vision in $K^bD^b-/-$ mice is also normal¹⁴. Thus, synapse elimination and eye-specific segregation fail to occur despite intact retinal activity patterns in $K^bD^b-/-$ mice, indicating that one or both of these MHC I proteins acts downstream of activity to drive synapse remodelling.

Neuronal H2-D^b rescues elimination and segregation

H2-D^b and H2-K^b are also critical for immune function and CD8 T-cell development¹⁹. Both MHC I molecules are expressed in LGN during the period of retinogeniculate synaptic refinement, with H2-D^b expressed at a higher level than H2-K^b (refs 10, 14). To separate a contribution of the immune system, and to examine if neuronal expression is sufficient for synapse elimination, H2-D^b expression was restored exclusively to neurons by crossing $K^bD^b-/-$ mice to $NSED^{b+}$ mice in which H2-D^b expression is regulated under the neuron-specific enolase (NSE) promoter²⁰. ‘Rescued’ offspring littermates have H2-D^b expression restored to CNS neurons while the rest of the body remains $K^bD^b-/-$ ($K^bD^b-/-$;

$n = 12/N = 6$); $K^bD^b-/-$ ($n = 23/N = 8$), $*P < 0.05$. **e**, Fibre fraction for wild type ($n = 12/N = 6$); $K^bD^b-/-$ ($n = 21/N = 8$), $**P < 0.01$, t -test for **d**, **e**, **f**, **g**, Intact retinal waves in $K^bD^b-/-$ at P10–P12. **f**, Raster plots of single-unit spike trains recorded from 10 representative RGCs during retinal waves. **g**, Correlation indices versus inter-electrode distance for all cell pairs for wild type ($N = 5$) versus $K^bD^b-/-$ ($N = 6$). Data correspond to mean values of medians from individual data sets and error bars represent s.e.m. $n = \text{cells}/N = \text{animals}$.

$NSED^{b+}$); ‘control’ littermates ($K^bD^b-/-$; $NSED^{b-}$) lack H2-K^b and H2-D^b throughout the body (Extended Data Fig. 3a). Genomic rescue, as well as low but highly significant levels of H2-D^b mRNA ($P = 0.0001$) and protein can be detected in $K^bD^b-/-$; $NSED^{b+}$ thalamus at P10 (Extended Data Fig. 3b–e). In contrast, no H2-D^b can be detected in spleen, gut or liver, with little if any expression in retina, hippocampus and cortex of $K^bD^b-/-$; $NSED^{b+}$ mice.

In $K^bD^b-/-$; $NSED^{b+}$ LGN neurons, only 1–2 EPSC steps could be evoked in response to increasing optic tract stimulus intensity (Fig. 2a), similar to the mature wild-type innervation pattern (see Fig. 1b) but very different from littermate $K^bD^b-/-$; $NSED^{b-}$ controls (Fig. 2b). Minimal stimulation also revealed an increase in SF-AMPA strength (Fig. 2c and Extended Data Figs 1c and 4b). Max-AMPA is similar between these genotypes (Extended Data Fig. 4a); thus fibre fraction in $K^bD^b-/-$; $NSED^{b+}$ LGN neurons is 56%, versus 25% in $K^bD^b-/-$; $NSED^{b-}$ neurons (Fig. 2d)—also markedly similar to wild type fibre fraction (see Fig. 1e). Thus, expression of H2-D^b in neurons rescues RGC synapse elimination in LGN of $K^bD^b-/-$ mice close to wild-type levels.

The formation of the adult anatomical pattern of eye-specific segregation in the LGN involves synapse elimination: initially intermixed retinal ganglion cell axons from the right and left eyes remodel, eventually restricting their terminal arborizations to the appropriate LGN layer^{21,22}. To examine whether eye-specific segregation in the LGN is also rescued, anatomical tract tracing methods^{14,23} were used at P34, an age chosen because it is more than 3 weeks after segregation is normally complete as assessed anatomically. The retinogeniculate projections in LGN of $K^bD^b-/-$; $NSED^{b+}$ mice appear almost indistinguishable from wild type, both in eye-specific pattern (Fig. 2e) and in per cent ipsi-contra overlap (Fig. 2f and Extended Data Fig. 5a, b). Segregation is impaired in control $K^bD^b-/-$; $NSED^{b-}$ littermates (Fig. 2e), as expected from previous studies of $K^bD^b-/-$ mice¹⁴. These anatomical results support the electrophysiological studies above and strongly suggest that both RGC synapse elimination and eye-specific segregation require neuronal H2-D^b.

Impaired LTD with natural activity patterns

Synapse elimination is thought to involve cellular processes leading to synaptic weakening such as LTD^{24,25}; conversely LTP-like mechanisms

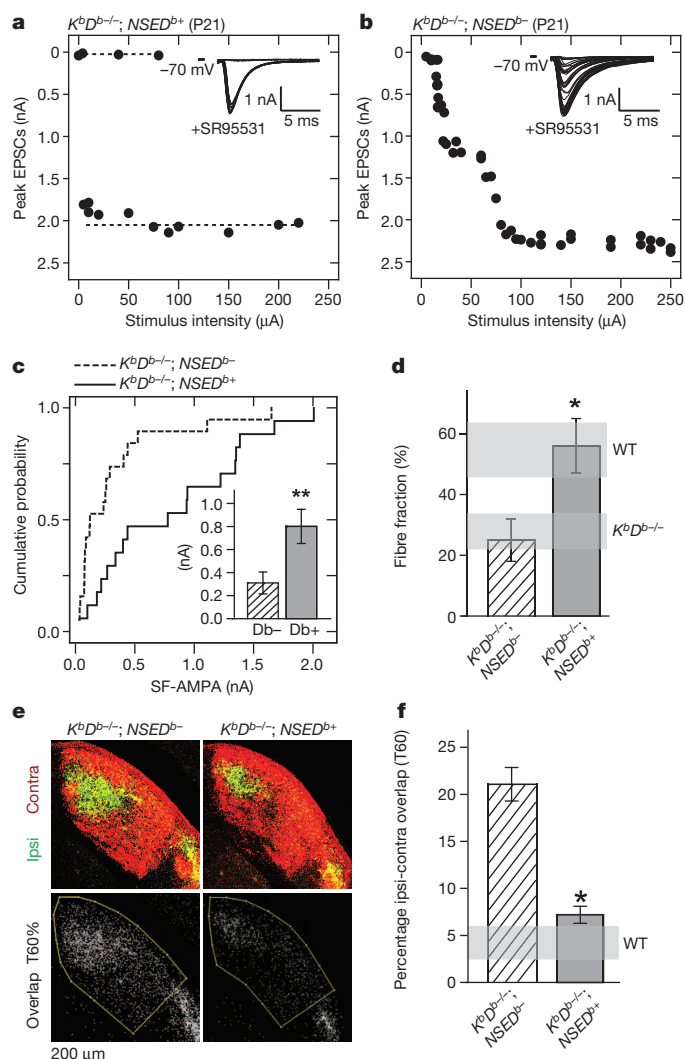


Figure 2 | H2-D^b expression in neurons rescues synapse elimination and eye-specific segregation in $K^bD^{b-/-}$ LGN. a–d, Rescue of synapse elimination at P20–P24. **a, b,** EPSC amplitudes versus optic tract stimulus intensity. Insets: example traces. **c,** Cumulative probability histogram of SF-AMPA. Inset: mean ± s.e.m. for control: D^{b-} ($K^bD^{b-/-}; NSED^{b-}$, $n = 19/N = 5$). Rescue: D^{b+} ($K^bD^{b-/-}; NSED^{b+}$, $n = 17/N = 7$). ** $P < 0.01$. **d,** Fibre fraction is also rescued in $K^bD^{b-/-}; NSED^{b+}$ ($n = 16/N = 7$) compared to $K^bD^{b-/-}; NSED^{b-}$ ($n = 18/N = 5$). * $P < 0.05$, Mann-Whitney U-test for c, d. Horizontal grey bars delineate Fig. 1e data (mean ± s.e.m.). **e, f,** Rescue of eye-specific segregation in $K^bD^{b-/-}; NSED^{b+}$ at P34. **e,** Top: coronal sections of dLGN showing pattern of retinogeniculate projections from the ipsilateral (green) and contralateral (red) eyes. Bottom: region of ipsi-contra pixel (white) overlap between the two channels at 60% intensity threshold (T60%). **f,** Percentage of dLGN area occupied by ipsi-contra overlap. mean ± s.e.m. for $K^bD^{b-/-}; NSED^{b-}$ ($N = 3$) and $K^bD^{b-/-}; NSED^{b+}$ ($N = 4$) (T60%) (* $P < 0.05$, two way ANOVA) (see Extended Data Fig. 5). Horizontal grey bar indicates wild-type value at T60% (from ref. 14). $n = \text{cells}/N = \text{animals}$.

are postulated for synaptic strengthening and stabilization^{26,27}. In addition, spike-timing-dependent mechanisms are crucial in *Xenopus* tectum for visually driven tuning of receptive fields²⁸. In mammalian LGN, LTP²⁹ or LTD³⁰ can be induced at retinogeniculate synapses using 100 Hz optic tract stimulation, which is far from the endogenous bursting patterns generated by retinal waves (Fig. 1f, g)^{1–4}. However, realistic patterns of optic tract stimulation mimicking waves, paired with postsynaptic depolarization of LGN neurons, have also been used; results revealed a synaptic learning rule that generates LTP when pre- and postsynaptic activity are coincident^{31,32}, but LTD when presynaptic optic tract activity precedes postsynaptic LGN depolarization within a broad window

corresponding to the 60–90 s duty cycle of retinal waves (Fig. 3a–c and Extended Data Fig. 2b)³¹. Moreover, using these timing patterns in conjunction with optogenetic stimulation of retina is sufficient either to drive or prevent segregation of RGC axons depending on the pattern³³. To determine whether synaptic learning rules based on natural activity patterns are altered at $K^bD^{b-/-}$ retinogeniculate synapses, perforated patch recordings were made in LGN slices from wild type versus $K^bD^{b-/-}$ at P8–P13, the relevant period when extensive synapse elimination and eye-specific segregation are actually occurring. First, paired-pulse stimulation was used to examine release probability: the same amount of synaptic depression was observed in wild-type and $K^bD^{b-/-}$ mice, indicating similar probabilities (Extended Data Fig. 6). Next, synchronous activity patterns were used, in which 10 Hz optic tract stimulation was paired with LGN depolarization (Fig. 3a, b: 0 ms latency), generating 10–20 Hz bursts of action potentials in LGN neurons mimicking retinal waves^{4,5}. In wild type, synchronous stimulation induced LTP (Fig. 3d, f; $117 \pm 8\%$ over baseline; $P < 0.001$). In $K^bD^{b-/-}$ LGN neurons, the same protocol elicited LTP indistinguishable from wild type (Fig. 3e, f).

In contrast, induction using asynchronous activity patterns reveals a defect in LTD. In wild type, when optic tract stimulation precedes LGN neuron depolarization by 1.1 s (Fig. 3a, c: 1,100 ms latency), LTD results (Fig. 3g, i: 12% decrease from baseline; $P < 0.001$). In contrast, in $K^bD^{b-/-}$, the same induction protocol failed to induce synaptic depression; if anything, a slight but significant potentiation was seen (Fig. 3h, i: 5% increase from baseline; $P < 0.005$). Thus, whereas LTD using asynchronous pre- and postsynaptic activity patterns is a robust feature of wild-type retinogeniculate synapses during the period of synapse elimination and eye-specific layer formation, it seems to be absent in $K^bD^{b-/-}$ mice. This impairment is consistent with the failure of synapse elimination and axonal remodelling in $K^bD^{b-/-}$ mice.

Ca²⁺-permeable AMPA receptors at $K^bD^{b-/-}$ synapses

Impaired LTD in $K^bD^{b-/-}$ mice could be due to altered regulation of NMDA (N-methyl-D-aspartate)-receptor-mediated synaptic responses, as LTP and LTD are known to be dependent on NMDA receptors at a variety of synapses²⁶. Surprisingly, the NMDA/AMPA ratio was not different between genotypes (Extended Data Fig. 7a, b). However, the kinetics of I_{AMPA} recorded in $K^bD^{b-/-}$ LGN neurons are markedly prolonged compared to wild type (Fig. 4a–d). The slowed decay in $K^bD^{b-/-}$ EPSCs is unlikely to be due to different peak I_{AMPA} amplitudes ($P > 0.1$; Fig. 4d), but could occur if there were greater Ca²⁺ influx through AMPA receptors.

Ca²⁺-permeable AMPA (CP-AMPA) receptors are blocked selectively by bath-applying the specific antagonist NASPM, a synthetic homologue of joro spider toxin³⁴. Indeed in $K^bD^{b-/-}$ LGN neurons, 100 μM NASPM blocked 40% of the current recorded at −70 mV but only 20% in wild type (Fig. 4e and Extended Data Fig. 7c), confirming a two-fold increase in CP-AMPA receptor-mediated currents in $K^bD^{b-/-}$ neurons. Another diagnostic feature of CP-AMPA receptors is rectification in the current–voltage (*I*–*V*) relationship when spermine is present in the internal recording solution^{34–36}. In wild type, the *I*–*V* relationship is linear. However in $K^bD^{b-/-}$ LGN neurons, rectification is very prominent (Fig. 4f, g and Extended Data Fig. 7d) but can be linearized close to wild-type levels by bath application of NASPM (Fig. 4f, g), indicating that the prominent *I*–*V* rectification in $K^bD^{b-/-}$ arises from an increase in CP-AMPA receptors.

Differences in composition of GluR subunits are known to modulate AMPA receptor Ca²⁺ permeability, and tetramers containing GluR2 confer Ca²⁺ impermeability³⁵. Indeed, the ratio of GluR1 to GluR2, the most prevalent subunits^{35,37,38}, is slightly increased in developing thalamus from $K^bD^{b-/-}$ mice (30% increase in $K^bD^{b-/-}$, $P = 0.07$; Extended Data Fig. 7e). The thalamus is highly heterogeneous, so we also examined cortical neuronal cultures: the ratio of GluR1 to GluR2 is also significantly increased (230% increase in $K^bD^{b-/-}$; $P = 0.03$; Extended Data Fig. 7f). Elevated levels of GluR1 subunits suggest that AMPA receptors in $K^bD^{b-/-}$ are more likely to be composed of GluR1 homomers,

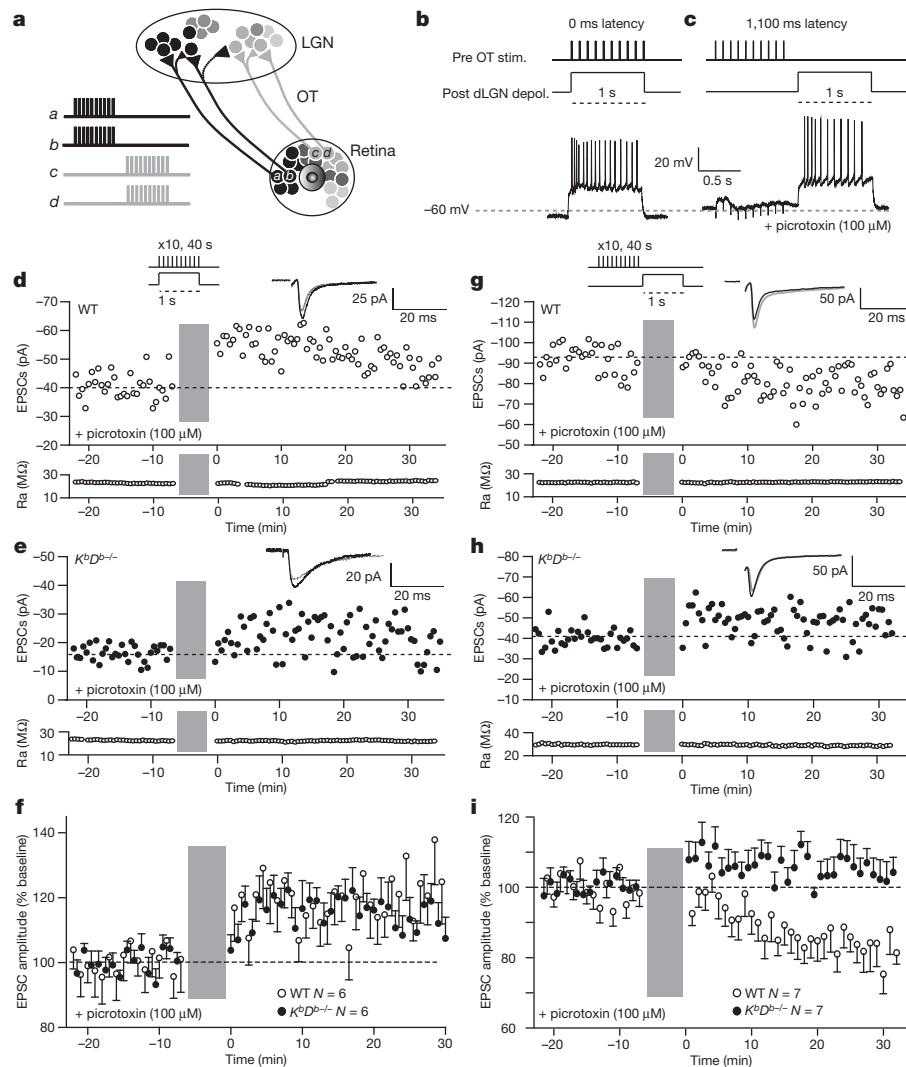


Figure 3 | Impaired LTD but intact LTP at retinogeniculate synapses in $K^bD^{b-/-}$ induced with natural activity patterns. **a**, Diagram illustrating basis for timing-dependent plasticity at developing retinogeniculate synapses. Spontaneous retinal waves propagate from 'a' towards 'c, d'; neighbouring RGCs fire synchronously but asynchronously with respect to RGCs located elsewhere. Waves drive action potentials in postsynaptic LGN neurons with varying time delays between pre- and postsynaptic activity. Ages P8–13 were studied. **b, c**, Top: conditioning protocol for LTP (0 ms latency; **b**) or LTD (1,100 ms latency; **c**). Bottom: example membrane potential changes recorded in LGN neuron during conditioning protocol. OT, optic tract. **d–f**, Intact LTP

yielding increased Ca^{2+} permeability. Together, results point to an increase in CP-AMPA receptors in $K^bD^{b-/-}$ mice. Similar increases in CP-AMPA receptors at other synapses are known to shift synaptic learning rules away from LTD and towards LTP^{39,40}. If so, the deficit in LTD observed with the asynchronous pairing protocol (Fig. 3c, i) in $K^bD^{b-/-}$ LGN should be rescued using NASPM to block CP-AMPA receptors—just what is observed (Extended Data Fig. 8).

Neuronal H2-D^b rescues synaptic function and LTD

If H2-D^b affects synapse elimination by regulating properties of AMPA receptors, then retinogeniculate EPSCs should be rescued to wild type in the LGN of $K^bD^{b-/-};\text{NSED}^{b+}$ mice. Indeed, the kinetics of I_{AMPA} are significantly faster in $K^bD^{b-/-};\text{NSED}^{b+}$ LGN neurons compared to $K^bD^{b-/-};\text{NSED}^{b-}$ LGN neurons (Fig. 5a–c and Extended Data Fig. 9a), indicating a decrease in Ca^{2+} -permeable AMPA receptors. Accordingly, NASPM-dependent inhibition of I_{AMPA} is only 20% in $K^bD^{b-/-};\text{NSED}^{b+}$, significantly reduced from the 40% inhibition observed in littermate

in $K^bD^{b-/-}$. Single experiment showing LTP in wild type (**d**) and $K^bD^{b-/-}$ (**e**). EPSC peak amplitude versus time. **f**, Summary of all 0 ms latency experiments: EPSC peak amplitude (% change from baseline) versus time ($n = 6/N = 6$ for each; $P > 0.1$, t -test). **g–i**, Deficient LTD in $K^bD^{b-/-}$. Single experiment for wild type (**g**) and $K^bD^{b-/-}$ (**h**). EPSC peak amplitude versus time. **i**, Summary of all 1,100 ms latency experiments: EPSC peak amplitude (% change from baseline) versus time ($n = 7/N = 7$ for each; $P < 0.01$, t -test). Grey bars indicate induction period. Insets: average EPSCs (30 traces) before (grey) and after (black) induction. **f, i**, 1 min data binning. Ra, access resistance (M Ω). $n = \text{cells}/N = \text{animals}$.

$K^bD^{b-/-};\text{NSED}^{b-}$ mice (Fig. 5d and Extended Data Fig. 9a). Moreover, the I - V relationship is linearized in $K^bD^{b-/-};\text{NSED}^{b+}$ LGN neurons when spermine is present in the internal recording solution, and bath application of NASPM has little additional effect ($P > 0.5$ at +40 mV), similar to wild type (Fig. 5e and Extended Data Fig. 9b). Because the Ca^{2+} permeability of AMPA receptors is close to wild-type levels in $K^bD^{b-/-};\text{NSED}^{b+}$ LGN, it is possible that LTD is also rescued. Indeed, the same asynchronous activity pattern that failed to induce LTD in $K^bD^{b-/-}$ (Fig. 3) induces robust LTD (15%; $P < 0.001$) in $K^bD^{b-/-};\text{NSED}^{b+}$ neurons, similar to wild type (Fig. 5f, g). Together, these observations suggest that restoring expression of H2-D^b in neurons is sufficient to rescue LTD at retinogeniculate synapses via decreasing the Ca^{2+} permeability of AMPA receptors.

Discussion

A major finding of this study is that the link between activity-dependent synapse pruning during development, and regulation of LTD and CP-AMPA

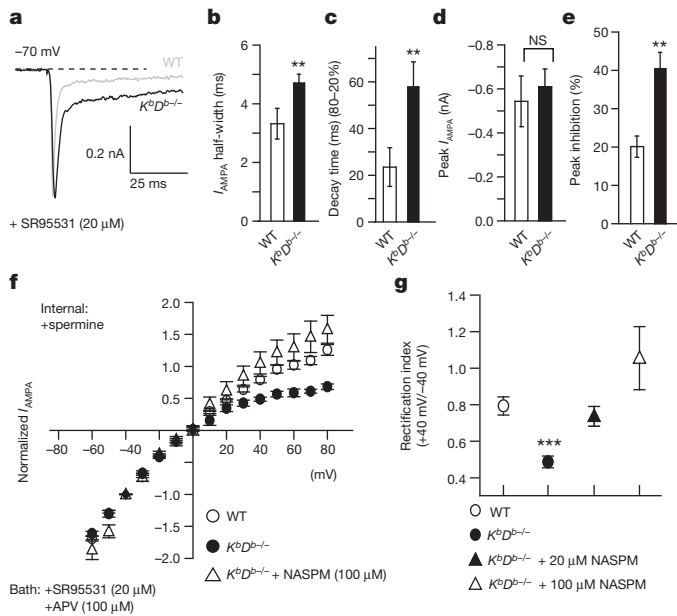


Figure 4 | Increased Ca^{2+} -permeable AMPA receptors at retinogeniculate synapses in $K^{bd-/-}$ LGN. **a–d**, Prolonged decay kinetics of I_{AMPA} in $K^{bd-/-}$ mice. **a**, Average I_{AMPA} (5–10 EPSCs) for wild-type versus $K^{bd-/-}$ LGN neurons. I_{AMPA} half-width (ms) (**b**), I_{AMPA} decay time (ms) (**c**) and peak amplitude (nA) (**d**) for wild type versus $K^{bd-/-}$ (WT, $n = 16/N = 4$; $K^{bd-/-}$, $n = 22/N = 5$). **e**, Increased per cent inhibition of peak I_{AMPA} by NASPM (100 μ M) in $K^{bd-/-}$ ($n = 13/N = 4$) versus wild type ($n = 9/N = 3$) (** $P < 0.01$; NS, not significant; Mann–Whitney U -test for **b**, **c**). **f**, I_{AMPA} I - V curves (normalized to -40 mV). **g**, Rectification index for wild type ($n = 14/N = 3$), $K^{bd-/-}$ ($n = 9/N = 3$), $K^{bd-/-}$ + 20 μ M NASPM ($n = 16/N = 4$) or $K^{bd-/-}$ + 100 μ M NASPM ($n = 6/N = 2$) (*** $P < 0.001$ for wild type versus $K^{bd-/-}$; $P > 0.05$ for wild type versus $K^{bd-/-}$ + NASPM (20 or 100 μ M), Mann–Whitney U -test). Ages studied: P8–P13. Also, see Extended Data Fig. 7. $n = \text{cells}/N = \text{animals}$.

receptors, requires neuronal MHCI function. It is notable that synapse elimination fails despite the fact that retinal waves and retinogeniculate basal synaptic transmission are intact. The persistence of multiple innervation in $K^{bd-/-}$ LGN neurons is highly reminiscent of the immature

synaptic connectivity in LGN of younger wild-type mice¹⁵, as well as of abnormal connectivity observed in LGNs of dark-reared or tetrodotoxin-treated wild-type mice¹⁶. Together, these considerations indicate that H2-D^b and H2-K^b act downstream of neural activity. In studying synaptic plasticity at retinogeniculate synapses, we imposed plasticity induction protocols that mimic natural patterns of spiking activity present in the retinogeniculate system during synapse elimination and eye-specific segregation. Our observation that in $K^{bd-/-}$ mice LTD is impaired while LTP is intact can explain the failure in retinogeniculate synapse elimination: if synapses cannot undergo weakening, then they cannot be eliminated. Because immunostaining for MHCI proteins H2-D^b and H2-K^b is co-localized with synaptic markers in array tomography¹⁴ and at synapses in immuno-electron microscopy^{14,41}, these observations also argue strongly that H2-D^b and/or H2-K^b at synapses regulate mechanisms of LTD, which in turn are required for synapse elimination. It would be useful to know whether other molecules implicated in RGC synapse elimination such as C1q (ref. 9), which co-localizes with H2-D^b and H2-K^b at synapses¹⁴, also alter LTD or instead act downstream of MHCI to target already weakened synapses for removal.

The rescue experiments performed here indicate that a single MHCI molecule—H2-D^b, when expressed in neurons—is sufficient for functional synapse elimination and anatomical eye-specific segregation in the LGN. By crossing $K^{bd-/-}$ mice to $NSED^b$ transgenic mice, expression of H2-D^b alone was restored to neurons but not elsewhere in the body, rescuing LTD, functional synapse elimination, Ca^{2+} -impermeable AMPA receptors and structural remodelling at retinogeniculate synapses. Notably, these brain phenotypes are rescued even though the immune system is still impaired in $K^{bd-/-};NSED^{b+}$ mice. Until this experiment, it was not known if any one MHCI molecule is sufficient either *in vitro* or *in vivo*, nor has it been possible to separate the general effects of immune compromise from the absence of H2-D^b and/or H2-K^b in neurons. Together, our observations argue for a key role for H2-D^b in reading out endogenous activity patterns into a lasting structural framework. In the human genome, as in mice, the MHCI (HLA) locus is large and highly polymorphic. Recent genome-wide association studies have consistently linked specific single nucleotide polymorphisms in MHCI to schizophrenia^{42,43}. Our observations offer possible mechanistic insight: alterations in expression levels of specific MHCI molecules at neuronal synapses could trigger changes in activity-dependent plasticity and synaptic pruning during critical periods of human development, generating lasting alterations in circuits and behaviour.

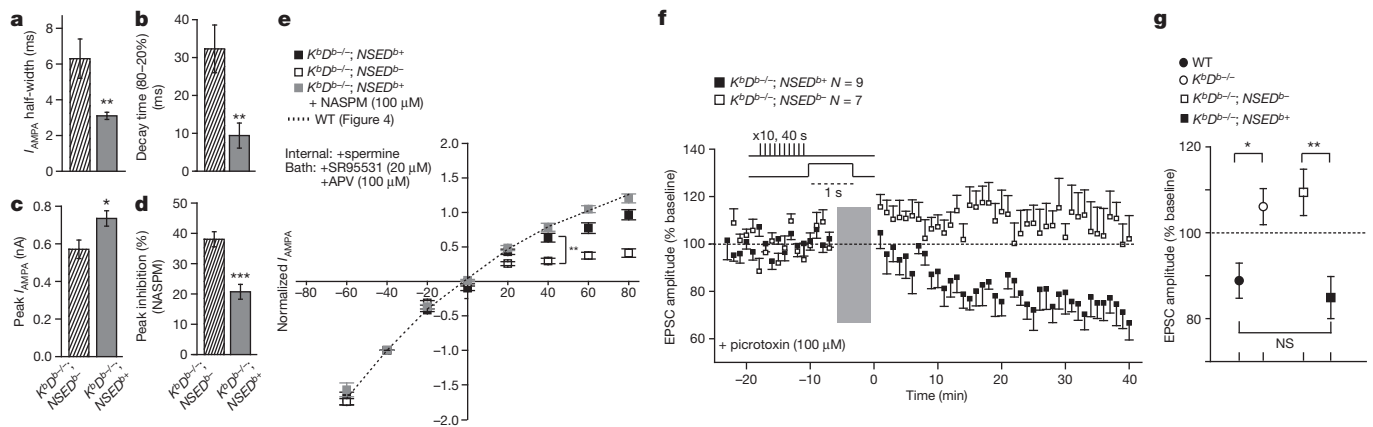


Figure 5 | Neuronal expression of H2-D^b restores Ca^{2+} -impermeable AMPA receptors and rescues LTD. **a–c**, I_{AMPA} half-width (ms) (**a**), I_{AMPA} decay time (ms) (**b**), and peak amplitude (nA) (**c**) for $K^{bd-/-};NSED^{b-}$ ($n = 9/N = 2$) and $K^{bd-/-};NSED^{b+}$ ($n = 11/N = 4$). **d**, Reduced per cent inhibition of peak I_{AMPA} by NASPM (100 μ M) in $K^{bd-/-};NSED^{b+}$ ($n = 10/N = 3$) compared to $K^{bd-/-};NSED^{b-}$ ($n = 8/N = 2$); * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$, Mann–Whitney U -test for **a–d**. **e**, Rescue of I_{AMPA} linear I - V relationship in $K^{bd-/-};NSED^{b+}$ LGN. Rectification index at +40 mV for $K^{bd-/-};NSED^{b-}$ ($n = 11/N = 3$) and $K^{bd-/-};NSED^{b+}$

($n = 13/N = 5$) shows significant difference (** $P < 0.005$); $K^{bd-/-};NSED^{b+}$ (+NASPM) ($n = 7/N = 3$) is not significantly different to $K^{bd-/-};NSED^{b+}$ ($P > 0.05$), Mann–Whitney U -test. See also Extended Data Fig. 9. mean \pm s.e.m. **f**, **g**, LTD rescued in $K^{bd-/-};NSED^{b+}$ LGN neurons. **f**, Ensemble average of all experiments at P8–9 (see Fig. 3). Grey bar indicates LTD induction period. 1 min data binning. **g**, Average per cent change (mean \pm s.e.m.) for wild type ($N = 7$), $K^{bd-/-}$ ($N = 7$), $K^{bd-/-};NSED^{b-}$ ($N = 7$) and $K^{bd-/-};NSED^{b+}$ ($N = 9$). * $P < 0.05$, ** $P < 0.01$, NS, not significant, t -test. P8–13 ages studied. $n = \text{cells}/N = \text{animals}$.

METHODS SUMMARY

All experimental protocols were approved by Stanford University Animal Care and Use Committees. $K^bD^{b-/-}$ mice were provided by H. Ploegh¹⁹ and $NSED^{b+/+}$ mice by M. B. A. Oldstone²⁰. These mice were maintained on C57BL/6 backgrounds. Crosses of these two lines generated $K^bD^{b-/-};NSED^{b+/+}$ mice plus littermate controls. Electrophysiological recordings were made from LGN neurons by cutting parasagittal brain slices containing dorsal lateral geniculate nucleus (dLGN) and optic tract; synaptic transmission and degree of innervation was assessed as previously described¹⁵. For plasticity experiments at retinogeniculate synapses, perforated patch-clamp technique and induction protocols with natural activity patterns were used³¹. Multi-electrode array recordings of retinal waves and anatomical labelling of retinogeniculate projections to determine status of eye-specific segregation were carried out according to ref. 18 and ref. 14. Pharmacological investigation of CP-AMPA receptors was carried out according to ref. 34. All experiments were conducted and analysed blind to genotype except in Fig. 4 (genotype was obvious to experimenter because of phenotype). Sample sizes were chosen for each experiment to reach statistical significance (P value equal to or less than 0.05; details given in Methods and figure legends).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 February 2013; accepted 13 February 2014.

Published online 30 March 2014.

- Meister, M., Wong, R. O., Baylor, D. A. & Shatz, C. J. Synchronous bursts of action potentials in ganglion cells of the developing mammalian retina. *Science* **252**, 939–943 (1991).
- Feller, M. B., Wellis, D. P., Stellwagen, D., Werblin, F. S. & Shatz, C. J. Requirement for cholinergic synaptic transmission in the propagation of spontaneous retinal waves. *Science* **272**, 1182–1187 (1996).
- Wong, R. O., Meister, M. & Shatz, C. J. Transient period of correlated bursting activity during development of the mammalian retina. *Neuron* **11**, 923–938 (1993).
- Mooney, R., Penn, A. A., Gallego, R. & Shatz, C. J. Thalamic relay of spontaneous retinal activity prior to vision. *Neuron* **17**, 863–874 (1996).
- Weliky, M. & Katz, L. C. Correlational structure of spontaneous neuronal activity in the developing lateral geniculate nucleus *in vivo*. *Science* **285**, 599–604 (1999).
- Ackman, J. B., Burbridge, T. J. & Crair, M. C. Retinal waves coordinate patterned activity throughout the developing visual system. *Nature* **490**, 219–225 (2012).
- Penn, A. A., Riquelme, P. A., Feller, M. B. & Shatz, C. J. Competition in retinogeniculate patterning driven by spontaneous activity. *Science* **279**, 2108–2112 (1998).
- Huberman, A. D., Feller, M. B. & Chapman, B. Mechanisms underlying development of visual maps and receptive fields. *Annu. Rev. Neurosci.* **31**, 479–509 (2008).
- Stevens, B. *et al.* The classical complement cascade mediates CNS synapse elimination. *Cell* **131**, 1164–1178 (2007).
- Huh, G. S. *et al.* Functional requirement for class I MHC in CNS development and plasticity. *Science* **290**, 2155–2159 (2000).
- Bjartmar, L. *et al.* Neuronal pentraxins mediate synaptic refinement in the developing visual system. *J. Neurosci.* **26**, 6269–6281 (2006).
- Coriveau, R. A., Huh, G. S. & Shatz, C. J. Regulation of class I MHC gene expression in the developing and mature CNS by neural activity. *Neuron* **21**, 505–520 (1998).
- Glynn, M. W. *et al.* MHCI negatively regulates synapse density during the establishment of cortical connections. *Nature Neurosci.* **14**, 442–451 (2011).
- Datwani, A. *et al.* Classical MHCI molecules regulate retinogeniculate refinement and limit ocular dominance plasticity. *Neuron* **64**, 463–470 (2009).
- Chen, C. & Regehr, W. G. Developmental remodeling of the retinogeniculate synapse. *Neuron* **28**, 955–966 (2000).
- Hooks, B. M. & Chen, C. Distinct roles for spontaneous and visual activity in remodeling of the retinogeniculate synapse. *Neuron* **52**, 281–291 (2006).
- Stevens, C. F. & Wang, Y. Changes in reliability of synaptic function as a mechanism for plasticity. *Nature* **371**, 704–707 (1994).
- Torborg, C. L., Hansen, K. A. & Feller, M. B. High frequency, synchronized bursting drives eye-specific segregation of retinogeniculate projections. *Nature Neurosci.* **8**, 72–78 (2005).
- Vugmeyster, Y. *et al.* Major histocompatibility complex (MHC) class I $K^bD^{b-/-}$ deficient mice possess functional CD8⁺ T cells and natural killer cells. *Proc. Natl Acad. Sci. USA* **95**, 12492–12497 (1998).
- Rall, G. F., Mucke, L. & Oldstone, M. B. Consequences of cytotoxic T lymphocyte interaction with major histocompatibility complex class I-expressing neurons *in vivo*. *J. Exp. Med.* **182**, 1201–1212 (1995).
- Shatz, C. J. & Kirkwood, P. A. Prenatal development of functional connections in the cat's retinogeniculate pathway. *J. Neurosci.* **4**, 1378–1397 (1984).
- Shatz, C. J. Emergence of order in visual system development. *Proc. Natl Acad. Sci. USA* **93**, 602–608 (1996).
- Torborg, C. L. & Feller, M. B. Unbiased analysis of bulk axonal segregation patterns. *J. Neurosci. Methods* **135**, 17–26 (2004).
- Zhou, Q., Homma, K. J. & Poo, M. M. Shrinkage of dendritic spines associated with long-term depression of hippocampal synapses. *Neuron* **44**, 749–757 (2004).
- Bastrikova, N., Gardner, G. A., Reece, J. M., Jeromin, A. & Dudek, S. M. Synapse elimination accompanies functional plasticity in hippocampal neurons. *Proc. Natl Acad. Sci. USA* **105**, 3123–3127 (2008).
- Malenka, R. C. & Bear, M. F. LTP and LTD: an embarrassment of riches. *Neuron* **44**, 5–21 (2004).
- Yuste, R. & Bonhoeffer, T. Morphological changes in dendritic spines associated with long-term synaptic plasticity. *Annu. Rev. Neurosci.* **24**, 1071–1089 (2001).
- Mu, Y. & Poo, M. M. Spike timing-dependent LTP/LTD mediates visual experience-dependent plasticity in a developing retinotectal system. *Neuron* **50**, 115–125 (2006).
- Mooney, R., Madison, D. V. & Shatz, C. J. Enhancement of transmission at the developing retinogeniculate synapse. *Neuron* **10**, 815–825 (1993).
- Ziburkus, J., Dilger, E. K., Lo, F. S. & Guido, W. LTD and LTP at the developing retinogeniculate synapse. *J. Neurophysiol.* **102**, 3082–3090 (2009).
- Butts, D. A., Kanold, P. O. & Shatz, C. J. A burst-based “Hebbian” learning rule at retinogeniculate synapses links retinal waves to activity-dependent refinement. *PLoS Biol.* **5**, e61 (2007).
- Shah, R. D. & Crair, M. C. Retinocollicular synapse maturation and plasticity are regulated by correlated retinal waves. *J. Neurosci.* **28**, 292–303 (2008).
- Zhang, J., Ackman, J. B., Xu, H. P. & Crair, M. C. Visual map development depends on the temporal pattern of binocular activity in mice. *Nature Neurosci.* **15**, 298–307 (2011).
- Liu, S. Q. & Cull-Candy, S. G. Synaptic activity at calcium-permeable AMPA receptors induces a switch in receptor subtype. *Nature* **405**, 454–458 (2000).
- Cull-Candy, S., Kelly, L. & Farrant, M. Regulation of Ca^{2+} -permeable AMPA receptors: synaptic plasticity and beyond. *Curr. Opin. Neurobiol.* **16**, 288–297 (2006).
- Isaac, J. T., Ashby, M. C. & McBain, C. J. The role of the GluR2 subunit in AMPA receptor function and synaptic plasticity. *Neuron* **54**, 859–871 (2007).
- Goel, A. *et al.* Cross-modal regulation of synaptic AMPA receptors in primary sensory cortices by visual experience. *Nature Neurosci.* **9**, 1001–1003 (2006).
- Hohnke, C. D., Oray, S. & Sur, M. Activity-dependent patterning of retinogeniculate axons proceeds with a constant contribution from AMPA and NMDA receptors. *J. Neurosci.* **20**, 8051–8060 (2000).
- Jia, Z. *et al.* Enhanced LTP in mice deficient in the AMPA receptor GluR2. *Neuron* **17**, 945–956 (1996).
- Toyoda, H. *et al.* Long-term depression requires postsynaptic AMPA GluR2 receptor in adult mouse cingulate cortex. *J. Cell. Physiol.* **211**, 336–343 (2007).
- Needleman, L. A., Liu, X. B., El-Sabeawy, F., Jones, E. G. & McAllister, A. K. MHC class I molecules are present both pre- and postsynaptically in the visual cortex during postnatal development and in adulthood. *Proc. Natl Acad. Sci. USA* **107**, 16999–17004 (2010).
- Stefansson, H. *et al.* Common variants conferring risk of schizophrenia. *Nature* **460**, 744–747 (2009).
- Ripke, S. *et al.* Genome-wide association analysis identifies 13 new risk loci for schizophrenia. *Nature Genet.* **45**, 1150–1159 (2013).

Acknowledgements We thank members of the Shatz laboratory for helpful comments. For technical assistance, we thank N. Sotelo-Kury, C. Chechelski and P. Kemper. For training in retinogeniculate slice methods, we thank C. Chen, P. Kanold and D. Butts. This work was supported by NIH grants R01 MH071666 and EY02858, and the G. Harold and Leila Y. Mathers Charitable Foundation (C.J.S.); NIH grant R01 EY13528 (M.B.F.); NDSEG and NSF Graduate Research Fellowships (J.D.A.); and an NSF Graduate Research Fellowship (L.A.K.).

Author Contributions H.L. and C.J.S. designed all experiments, analysed and reviewed all results and wrote manuscript. Data contributions are as follows: electrophysiology experiment by H.L.; multi-electrode array experiments by L.A.K. and M.B.F. B.K.B. designed H2-D^b monoclonal antibody and performed western blots. H.L. designed and performed RT-PCR experiments. A.D. performed RGC neuronal tract tracing experiments and analysis. J.D.A. and S.C. performed Taqman qPCR.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.J.S. (cshatz@stanford.edu).

METHODS

All experimental protocols were approved by Stanford University Animal Care and Use Committees. $K^bD^{b-/-}$ mice were provided by H. Ploegh¹⁹ and $NSED^b$ mice by M. B. A. Oldstone²⁰. These mice were maintained on C57BL/6 backgrounds. Crosses of these two lines generated $K^bD^{b-/-};NSED^{b+}$ mice plus littermate controls. Electrophysiological recordings were made from LGN neurons by cutting parasagittal brain slices containing dorsal lateral geniculate nucleus (dLGN) and optic tract; synaptic transmission and degree of innervation was assessed as previously described¹⁵. For plasticity experiments at retinogeniculate synapses, perforated patch-clamp technique and induction protocols with natural activity patterns were used³¹. Multi-electrode array recordings of retinal waves and anatomical labelling of retinogeniculate projections to determine status of eye-specific segregation were carried out according to ref. 18 and ref. 14. Pharmacological investigation of CP-AMPA receptors was carried out according to ref. 34. All experiments were conducted and analysed blind to genotype except in Fig. 4 (genotype was obvious to experimenter because of phenotype). Sample sizes were chosen for each experiment to reach statistical significance (P value equal to or less than 0.05; details given in Methods and figure legends).

Genotyping. For $K^bD^{b-/-}$: three primer sets were mixed together (A: 5'-CGG AAGTCGCCTTACCTGATTG-3', B: 5'-CAGCAGAAACATACAAGCTGTC-3', C: 5'-GTCTCTCTGGCACCCTATGGG-3'), from which bands of 520 bp were detected for wild type, 400 bp for $K^bD^{b-/-}$ and double bands of 520 bp + 400 bp for heterozygote $H2-K^{b+/-}H2-D^{b+/-}$. For $NSED^b$ mice and $K^bD^{b-/-};NSED^{b+}$: several primer sets were used, resulting in reliable and consistent results: a primer set consisting of $NSED^b$ -A (5'-GAGATCGACTCTAGAGGATC-3') and $NSED^b$ -B1 (5'-GCGCTCTGGTGTAGTACGC-3')²⁰ or $NSED^b$ -B2 (5'-GTAGCCGAGCA GGTCTCTCA-3') amplifies part of NSE promoter region and exon 2 of $H2-D^b$ cDNA (~500 bp) (Extended Data Fig. 3b). Another primer set was designed to detect a unique D^b minigene expression cassette of ~800 bp (A: 5'-CGACACAGG GACGCTGACG-3', B: 5'-CAGTCCAATGATGGCCATAG-3). Taqman qPCR primer mix ($H2-D1/H2-K1$ (Mm04208017_mH) was also used, with same results.

Slice preparation. Parasagittal brain slices containing the dorsal lateral geniculate nucleus (dLGN) and optic tract were obtained as previously described^{15,44}. The brain was removed rapidly and immersed in oxygenated ice-cold choline-based slicing solution (composition in mM: 78.3 NaCl, 23 NaHCO₃, 33.8 choline chloride, 2.3 KCl, 1.1 NaH₂PO₄, 6.4 MgCl₂·6H₂O, 0.45 CaCl₂·2H₂O, 23 glucose, aerated with 95% O₂/5% CO₂). The hemispheres were separated at an angle of 5–10° relative to midline. The larger hemisphere including optic tract was glued onto the cutting stage of a vibratome (Leica VT1000S) at an angle of 15–20°, which puts optic tract and LGN in the same plane relative to the blade; 250-μm-thick slices were cut. Note that only one slice per animal containing both intact optic tract and LGN can be obtained due to the small size of the LGN at relevant ages. Cortex was separated from thalamus to prevent excitatory recurrent disynaptic responses between dLGN and corticothalamic neurons. Slices were recovered at 31 °C for 30 min in choline chloride slicing solution, and then for 30 min in artificial cerebrospinal fluid (ACSF) (ACSF composition in mM: 125 NaCl, 26 NaHCO₃, 2.3 KCl, 1.26 KH₂PO₄, 1.3 MgSO₄·7H₂O, 2.5 CaCl₂·2H₂O, 25 glucose, aerated with 95% O₂/5% CO₂).

Electrophysiology. All recordings were at room temperature in a chamber with constant ACSF flow, except for paired-pulse ratio measurements, which were made at 30–32 °C. Stimulating electrode(s) were placed along the optic tract at the ventral end of dLGN and/or lateral edge of dLGN aligned with the recording site. dLGN neurons were distinguished by their large soma size and 3+ primary dendrites¹⁵. For synaptic current measurements, glass electrodes were filled with Cs⁺-based internal recording solution (2–4 MΩ, composition in mM: 105 CsCl, 20 TEA-Cl, 2 MgCl₂, 1 EGTA, 10 HEPES, 3 Mg-ATP, 15 phosphocreatine, 1 Na-GTP, 5 QX-314, pH 7.4, 280 mOsm). To block inhibitory synaptic responses, SR95531 (20 μM; Tocris) or picrotoxin (100 μM; Sigma-Aldrich) was bath applied. AMPA receptors were blocked with DNQX (20 μM; Tocris), and NMDA receptors were blocked with DL-APV (100 μM; Tocris). For NMDA/AMPA ratio and spermine (0.1 mM, Sigma-Aldrich) dependent I_{AMPA} measurements, D600 (methoxyverapamil hydrochloride, 0.1 mM; Tocris) was added in Cs⁺-based internal solution to block voltage-gated calcium channel activation. Spermine-containing internal solution was prepared fresh every 1–2 h. NASPM (1-naphthyl acetyl spermine trihydrochloride, 20–100 μM; Tocris) was bath applied. Rectification index was calculated by dividing peak I_{AMPA} at +40 mV with at –40 mV from $I-V$ curves measured using spermine-containing internal solution³⁴.

For patch-clamp recordings: synaptic responses were recorded using an Axopatch 200B or 700A amplifier (Axon Instruments/Molecular Devices), digitized using Digidata 1322A (Axon Instruments) and data acquisition was performed by Clampex 9.2 (Axon Instrument). For whole-cell voltage clamp experiments, the series resistance (R_s) was corrected by 60–80% after cancellation of capacitive components. For synaptic plasticity experiments in which EPSC size was less than 120 pA,

compensation was omitted. Calculated liquid junction potentials (Cs⁺ internal: 4 mV; K⁺ internal: 9 mV) were not compensated, except for spermine-dependent $I-V$ relationship plots (Figs 4 and 5). Series resistance (R_s) was constantly monitored throughout experiments by injecting small hyperpolarizing currents.

For synaptic plasticity experiments: perforated patch-clamp technique was performed using a K⁺-based internal solution containing amphotericin B (Fisher Scientific) (composition in mM: 115 K-methane Sulphonate, 20 KCl, 0.2 EGTA, 10 HEPES, 4 Mg-ATP, 10 phosphocreatine, 0.3 Na-GTP, pH 7.3 with KOH). Note that perforated patch recordings are necessary to prevent dialysis of intracellular components required for plasticity^{29,31,45}. Amphotericin B stock solution (40–60 mg ml⁻¹ in DMSO) was prepared fresh every day, from which 100–200 μg ml⁻¹ internal solution was prepared hourly and kept on ice. The tip of the patch electrode was filled with internal solution without amphotericin B before back-filling with amphotericin B containing internal solution. Following the formation of GΩ seal, gradual drop of access resistance was monitored until stabilizing at 20–120 MΩ. Trials with sudden changes in R_s (clearly discernible due to rupture of perforation or clogging of patch pipette by amphotericin B precipitation) were discarded. EPSC changes with stable R_s within ±20 MΩ were used for analysis except when gradual changes in R_s were correlated with changes in EPSCs size. After obtaining a stable baseline for 10–20 min, the configuration was switched to current-clamp mode and plasticity induction protocols were applied by pairing optic tract stimulation with postsynaptic current injection for 1 s to generate 10–20 Hz action potentials either synchronously or with 1,100 ms latency (Fig. 3a–c)³¹. V_m (resting membrane potential measured at $I = 0$ configuration in perforated patch clamp mode) was not different between wild type (–58.5 ± 1.2 mV; $N = 16$ cells from 16 animals), $K^bD^{b-/-}$ (–59.6 ± 1.3 mV; $N = 10$), $K^bD^{b-/-};NSED^{b+}$ (–60.5 ± 1.9 mV; $N = 13$) or $K^bD^{b-/-};NSED^{b-}$ (–62.2 ± 2.5 mV; $N = 7$) ($P > 0.1$, t -test). The pairing protocol was repeated 10 times over a 40 s interval. Data were acquired with resting membrane potentials (measured at $I = 0$) greater than –53 mV. For the NASPM LTD rescue experiments (Extended Data Fig. 8), 20 μM NASPM was added to the bath, and the asynchronous pairing protocol (10 times over a 40 s interval) was repeated twice, separated by 20 min. $N =$ one cell per each animal. Percentage change is mean ± s.e.m. of EPSC amplitude averaged over 30 min after conditioning period as compared to baseline per cell.

For optic tract stimulation, concentric bipolar (CBABD70; FHC Inc.) or low-resistance glass-electrode filled with ACSF were used for stimulating electrodes without noticeable differences. Stimulus strength was regulated using an ISO-Flex stimulus isolator (A.M.P.I.) and applied at 0.025–0.033 Hz. For plasticity experiments, stimulus strength was set to achieve stable EPSCs (10–120 pA) activated by a single or a few fibres, and stimuli were applied at 0.025 Hz. To measure spermine-dependent $I-V$ rectification, stimulation intensity was adjusted to evoke I_{AMPA} currents in the 100–300 pA range, due to the presence of Ca²⁺-permeable AMPA receptor-mediated synaptic responses in $K^bD^{b-/-}$ mice.

Evaluation of the number of synaptic inputs. The number of connected fibres was estimated using two approaches modified from ref. 15. First, synaptic responses were recorded while varying optic tract stimulus intensity from minimal to maximal. Stimulation in increments of 1.0–2.5 μA was used between 0 μA and 100 μA, and larger increments (10–20 μA) were used at stimulation intensities greater than 100 μA. The stimulus-response profile provides an estimate of the number of RGC inputs to the cell. We also estimated the number of inputs to each LGN neuron based on the single-fibre AMPA current elicited at –70 mV by minimal stimulation as a per cent of the total current elicited by maximal stimulation for the same cell (Fibre fraction). For minimal stimulation, the method is based on the assumption that minimal stimulation elicits synaptic responses from only a single fibre in the optic tract (SF-AMPA). After a synaptic response was observed, the intensity of electrical stimulation was lowered to obtain a failure rate higher than 50% (20–50 repeats). Synaptic responses with an onset latency of <5 ms within a 1-ms window were regarded as direct, monosynaptic responses. The mean peak amplitudes of successful responses were taken as the strength of the single fibre synaptic response (SF-AMAP). Onset latency of SF-AMPA by minimal stimulation (estimated as time to reach 10% of peak I_{AMPA}) was not different between wild type, $K^bD^{b-/-}$, $K^bD^{b-/-};NSED^{b-}$ or $K^bD^{b-/-};NSED^{b+}$ (Extended Data Fig. 1c). For maximal stimulation, electrical stimulation was increased until the peak synaptic currents reached steady-state level (1–600 μA range). Fibre fraction was calculated as a ratio of single-fibre synaptic strength (SF-AMPA) to maximal synaptic strength (Max-AMPA) from the same cell (FF = SF/Max-AMPA per cell, %). The number of cells recorded per animal was limited to 4 for SF-AMPA and Max-AMPA measurements.

Kinetics of EPSCs were measured as both the width at half-maximal EPSC amplitude (half-width), and decay time within 80–20% range of peak EPSCs using Clampfit.

Multi-electrode array recordings of retinal waves. Retinas were isolated from wild-type and $K^bD^{b-/-}$ mice and pieces of retina were placed retinal-ganglion-cell-side

down onto a 60-electrode multi-electrode array arranged in an 8×8 grid (excluding the four corners) with $10 \mu\text{m}$ diameter electrodes at $100 \mu\text{m}$ spacing (Multi Channel Systems)^{13,46}. The retina was held in place with a weighted piece of dialysis membrane and superfused continuously with artificial cerebral spinal fluid (ACSF) containing (in mM): 119.0 NaCl, 26.2 NaHCO_3 , 11 glucose, 2.5 KCl, 1.0 K_2HPO_4 , 2.5 CaCl_2 , and 1.3 MgCl_2 . ACSF was oxygenated with 95% O_2 and 5% CO_2 and maintained at 32°C , pH 7.4. Each preparation was allowed to equilibrate for 20 min before data acquisition. Spontaneous firing patterns were then recorded for 60 min. Voltage traces on each electrode were sampled at 20 kHz and filtered between 120 and 2,000 Hz. Events that crossed a spike threshold were sorted offline to identify single units using Plexon Offline Sorter software. Single units were identified by combining principal component analysis together with a valley-seeking algorithm and were inspected manually. The mean firing rate of all units throughout the recording was calculated and units with a mean firing rate less than 10% of overall mean firing rate were excluded from further analysis. Spike-sorted data were analysed in MATLAB (Mathworks). To identify bursts, a modified Poisson Surprise algorithm was used as described previously. Mean burst values for each retina were grouped according to age (P5–P8 and P10–P12) and genotype (wild type and $K^b\text{D}^{b-/-}$). Differences between means of different genotypes within each age group were evaluated for statistical significance using a *t*-test. To determine spatial properties of waves, a pair-wise correlation index, CI, was computed as a function of distance between two cells for all spikes in the recording, using a method described previously^{3,46}. The correlation index gives a measure of the likelihood relative to chance that a pair of neurons fire together within a 100-ms time window. The distance between cells was approximated as the distance between the electrodes on which the activity of the cells was recorded. The cell pairs were grouped according to their intercellular distance, and medians were computed over all cell pairs in a given distance group for each individual retina. The median correlation indices were then averaged across retinas, for each age and genotype, and plotted as a function of intercellular distance.

Anterograde labelling of retinal ganglion axons and multiple threshold analysis. All methods were performed according to previously described studies^{14,47}. P31–34 mice were anaesthetized with isoflurane. Cholera toxin B (CTB) subunit (1–2 μl) conjugated to AF488 was injected in the right eye and CTB subunit conjugated to AF594 was injected in the left eye (1 mg ml^{-1} dissolved in 0.2% DMSO, Invitrogen). After 24 h, animals were perfused and fixed by transcardial perfusion of 0.1 M phosphate buffered saline (PBS), then ice-cold 4% paraformaldehyde (PFA) in 0.1 M PBS. Brains were post-fixed in 4% PFA overnight. $100\text{-}\mu\text{m}$ -thick sections were cut on a freezing microtome. Sections were mounted with Prolong Antifade Gold (Invitrogen), and after coverslipping were imaged on a Zeiss Leica PS2 confocal microscope (Leica Microsystems). dLGN sections taken at the middle of the rostral-caudal extent of dLGN where the area of the ipsilateral eye projection is greatest were imaged in red and green channels. Images were acquired and analysed such that the peak intensity values were below maximum, and multiple threshold analysis was used by varying red and green channels at intensity thresholds of 20%, 40%, 60%, 80% and 100% of maximum. To measure the amount of overlap between inputs from both eyes, pixels overlapping in both red and green channels were identified using Image J (NIH) and the Colocalization plug-in tool. The total area of overlapping pixels was represented as a percentage of total dLGN area. Two-way ANOVA was used for statistical analysis.

RT-PCR. RNA was prepared from tissue using RNeasy RNeasy-4PCR (Ambion, Life Technologies) followed by cDNA synthesis using iScript Select cDNA synthesis kit (Bio-Rad). RT-PCR was performed using 0.5–1.0 μg of template cDNA using primers as follows (Extended Data Fig. 3b, c). H2-D^b: A, 5'-CAAGAGC AGTGGTTCGAGTGAG-3'; B, 5'-CTTGTAATGCTCTGCAGCACCACCT-3'. GAPDH (glyceraldehyde-3-phosphate dehydrogenase; used as a reference gene): A, 5'-ATTGTCAGCAATGCATCTCTGC-3'; B, 5'-AGACAACCTGGTCTCA GTGT-3'.

The quality of cDNAs was confirmed by genotyping PCR reactions using 0.5–1 μg of cDNA as template and the samples with genomic DNAs were discarded. H2-D^b-specific RT-PCR bands from $K^b\text{D}^{b-/-}$; NSD^{b+} were confirmed by sequencing following cloning into pCR2.1-TOPO TA vector (Life Technology Corporation). **Taqman qPCR.** RNA was extracted from each thalamus and cDNA was synthesized using iScript cDNA Synthesis kit (Bio-Rad). Gene expression was analysed with Taqman Gene Expression Assays (Applied Biosystems) for H2-D1/H2-K1 (Mm04208017_mH) and housekeeping gene GAPDH (Mm99999915_g1). All reaction mixes contained 5% $20\times$ Taqman Gene Expression Assay, 50% Taqman Master Mix (4304437), 25% MilliQ water, and 200 ng of cDNA in 20 μl total volume per well. Experiments were carried out on a 7300 Real Time PCR system (Applied Biosystems). The relative amount of tested message was normalized to GAPDH, and efficiency corrected based on standard curves. Gene expression levels were normalized relative to wild-type control, which was set to 1. All samples were run

in triplicate at least three times. Analyses were carried out as previously described using one-way ANOVA⁴⁸.

Immunoprecipitation/western blots of H2-D^b. Brain lysates in lysis buffer (150 mM NaCl, 50 mM Tris, 0.25% sodium deoxycholate, 1% NP-40, 1 mM EGTA, 1 mM PMSF, 1X Pefabloc (Roche)) were prepared by shearing ten times in a dounce homogenizer, then centrifuging at $12,000g$ for 10 min. Supernatants were precleared by incubation with Protein G-agarose beads (Invitrogen), and protein amounts were measured using a BioRad protein assay. Normalized lysates were then incubated with 12 μg anti-H2-D^b antibody 28-14-8 (BD Biosciences, 553600) overnight at 4°C . Protein G-agarose was incubated with samples for 45 min. Beads were washed three times with lysis buffer, and then heated to 85°C for 3 min in NuPAGE LDS $4\times$ sample buffer (Invitrogen) + 1% β -2 mercaptoethanol. Samples were then electrophoresed on an SDS-PAGE gel, transferred to Immobilon-P PVDF transfer membrane (Millipore), and western blotted with rabbit monoclonal antibodies to H2-D^b, made against the extracellular domain of H2-D^b.

GluR1/2 expression levels. Thalami were isolated from P22 wild-type or $K^b\text{D}^{b-/-}$ mice, and synaptosome-enriched fractions prepared as described^{49,50}. Freshly isolated individual thalami were homogenized in homogenization buffer (10 mM HEPES pH 7.3, 0.5 mM EGTA, 33% sucrose, 4 mM Pefabloc SC PLUS (Roche), and 0.2 mM phenylmethanesulphonyl fluoride (PMSF)), and centrifuged (10 min at $2,000g$). Supernatants were passed through three layers of $100 \mu\text{m}$ pore nylon membranes, and then through $5 \mu\text{m}$ nitrocellulose filters. Filtrates were centrifuged ($10,000g$ for 10 min), and pellets re-suspended in homogenization buffer. Protein amounts were determined by BioRad Protein Assay (BioRad) (about 200 μg synaptosomal protein was obtained per thalamus), and SDS-PAGE sample buffer (Novex) was added to a $1\times$ concentration. After heating (85°C for 4 min), equivalent amounts of protein from each sample (20 μg per sample) were electrophoresed on a SDS-PAGE gel, transferred to PVDF membrane (Bio-Rad), and western blotted with antibodies to GluR1 (Millipore, AB1504), GluR2 (Abcam, AB133477), N-tubulin (Abcam, AB18207), or GAPDH (Abcam, AB9485).

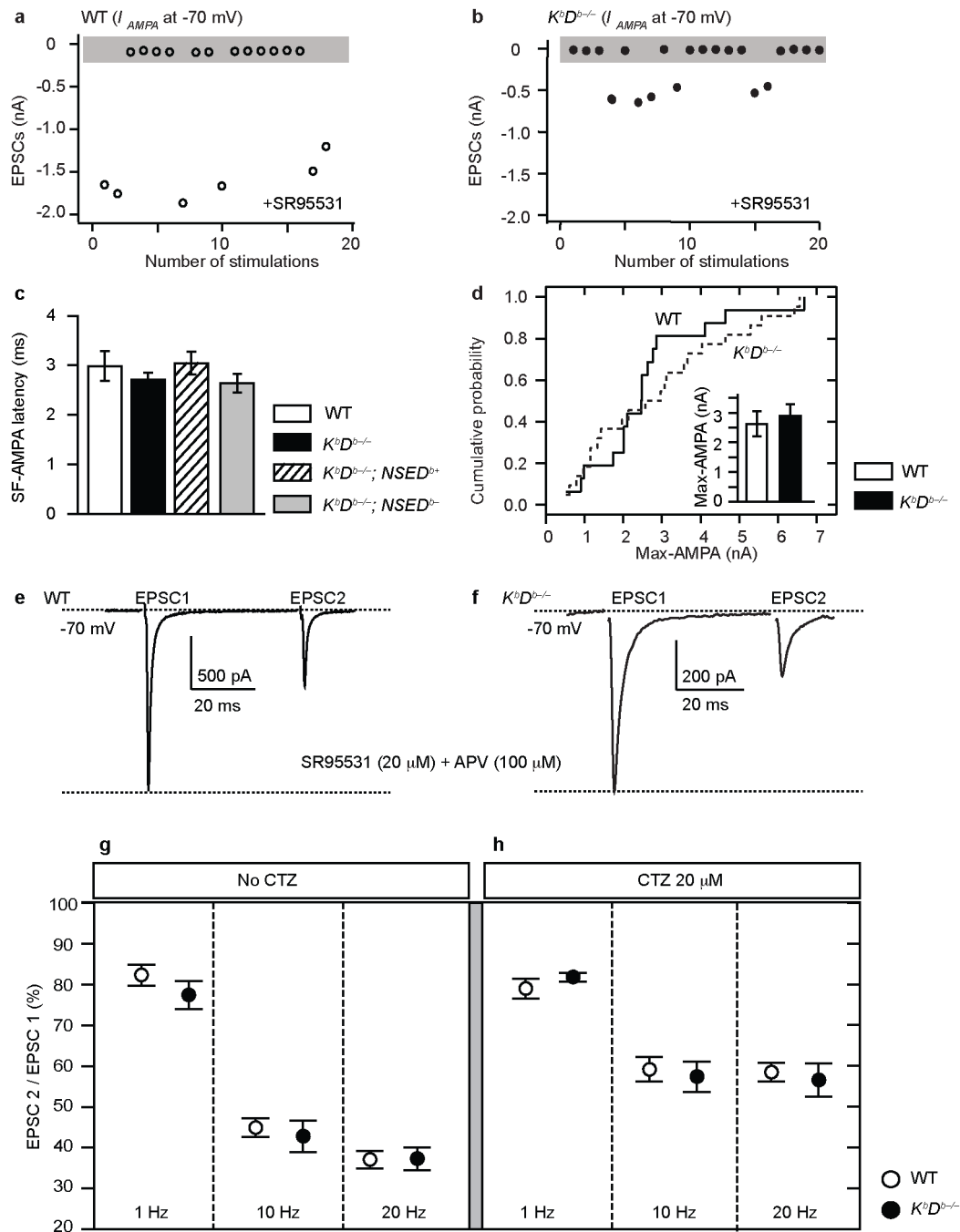
For quantification, gels were scanned and intensities of bands were quantified using the Image J program. After quantification, intensity values for GluR1 and GluR2 were normalized by dividing them against each GAPDH or N-tubulin values from the same thalamus. Western blots typically were run with multiple wild-type and $K^b\text{D}^{b-/-}$ samples; to compare results from different western blots, the intensity measurement for each sample thalamus was normalized against the average value derived from all wild-type bands on that same blot. The GluR1/GluR2 ratio for each thalamus was then determined. Finally, an average GluR1/GluR2 ratio was calculated from each individual thalami (Extended Data Fig. 7e).

Mouse cortical cultures were made as described⁵¹. Wild-type or $K^b\text{D}^{b-/-}$ cortical neurons were isolated at E16.5, and cultured on 6-cm tissue culture plates for 14 days *in vitro*. Plates were then washed twice with PBS and cells were lysed directly on the plates with lysis buffer (150 mM NaCl, 50 mM Tris-HCl pH 7.5, 0.25% sodium deoxycholate, 1% NP-40, 1 mM EGTA, 1 mM PMSF, 1X Pefabloc SC PLUS (Roche)) for 10 min. Lysates were then centrifuged at $12,000g$ for 10 min. Supernatants were normalized by BioRad Protein Assay (BioRad), and equivalent amounts of protein from each sample (25 μg per sample) were subjected to SDS-PAGE and western blotting with antibodies to GluR1, GluR2, GAPDH, and/or N-tubulin, as described above. Gels were scanned and bands quantified using Image J. Values were then normalized by dividing them by GAPDH or N-tubulin values, and an average GluR1/GluR2 ratio was calculated from each individual sample (Extended Data Fig. 7f). Non-parametric Mann–Whitney *U*-test (Prism software) was used to calculate *P* values for all analyses (two-tailed).

Data analysis and statistics for electrophysiology experiments. Data analysis as well as graph plotting were done using Clampfit 10.2 (Axon Instruments) and the OriginPro8.0 (OriginLab Corporation). Cages containing mice of different genotypes used for each experiment were relabelled and randomized so that experimenter did not know genotype. Experiments were then analysed blind to genotype unless stated in figure legends. In addition, Student's *t*-test was used for data with normal distribution and Mann–Whitney *U*-test for data with non-normal distribution. Shapiro–Wilk test was conducted for normality test. *N* = number of animals; *n* = number of cells. Average data are presented as mean \pm s.e.m., unless stated otherwise. For synaptic plasticity experiments, *t*-test was used. Two-tailed analysis was used throughout the study.

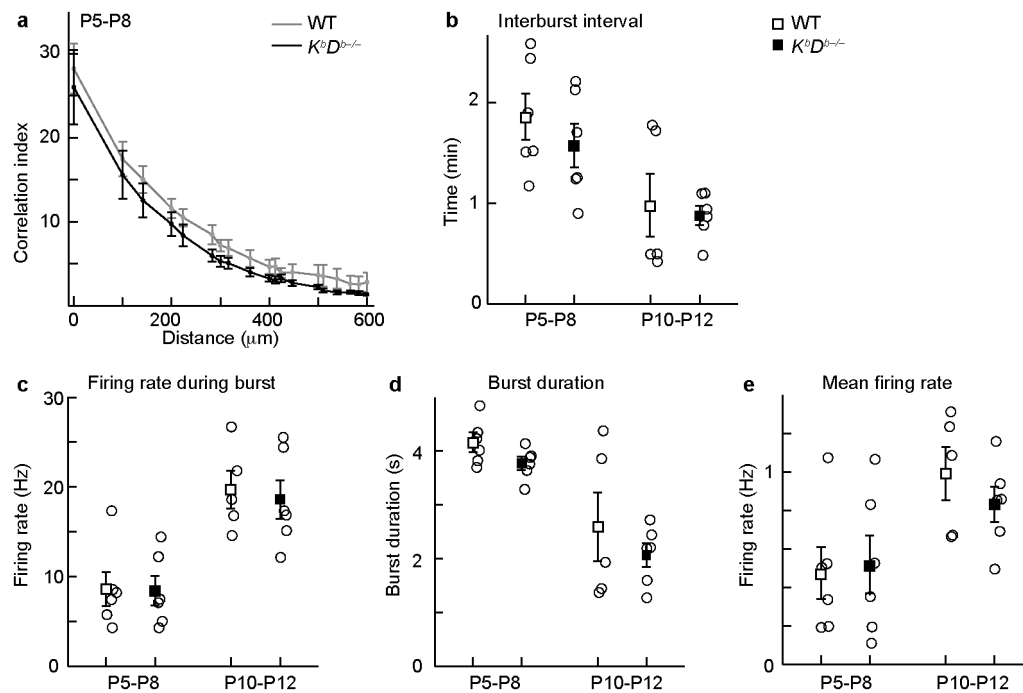
44. Turner, J. P. & Salt, T. E. Characterization of sensory and corticothalamic excitatory inputs to rat thalamocortical neurones *in vitro*. *J. Physiol. (Lond.)* **510**, 829–843 (1998).
45. Rae, J., Cooper, K., Gates, P. & Watsky, M. Low access resistance perforated patch recordings using amphotericin B. *J. Neurosci. Methods* **37**, 15–26 (1991).
46. Torborg, C. L., Hansen, K. A. & Feller, M. B. High frequency, synchronized bursting drives eye-specific segregation of retinogeniculate projections. *Nature Neurosci.* **8**, 72–78 (2005).
47. Torborg, C. L. & Feller, M. B. Unbiased analysis of bulk axonal segregation patterns. *J. Neurosci. Methods* **135**, 17–26 (2004).

48. Adelson, J. D. *et al.* Neuroprotection from stroke in the absence of MHC1 or PirB. *Neuron* **73**, 1100–1107 (2012).
49. Johnson, M. W., Chotiner, J. K. & Watson, J. B. Isolation and characterization of synaptoneurosomes from single rat hippocampal slices. *J. Neurosci. Methods* **77**, 151–156 (1997).
50. Yin, Y., Edelman, G. M. & Vanderklish, P. W. The brain-derived neurotrophic factor enhances synthesis of Arc in synaptoneurosomes. *Proc. Natl Acad. Sci. USA* **99**, 2368–2373 (2002).
51. Viesselmann, C., Ballweg, J., Lumbard, D. & Dent, E. W. Nucleofection and primary culture of embryonic mouse hippocampal and cortical neurons. *J. Vis. Exp.* (2011).



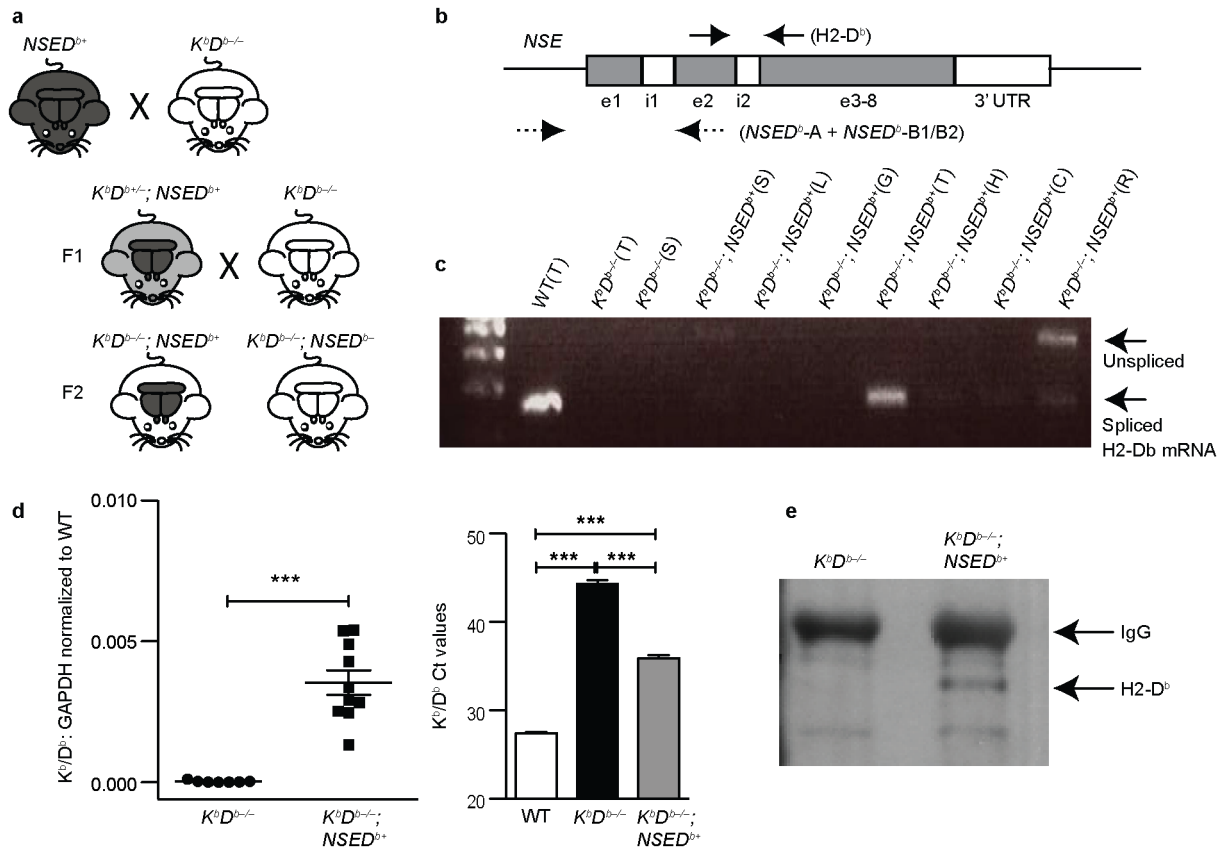
Extended Data Figure 1 | Comparison of retinogeniculate synaptic responses in wild type versus $K^bD^{b-/-}$. **a, b,** Examples of minimal stimulation for wild type (open circles) and $K^bD^{b-/-}$ (filled circles). Plot of EPSC peak versus number of stimulations (grey box represents failures, >50%). **c,** No difference in onset latency of SF-AMPA between all genotypes. Onset latency of SF-AMPA was estimated using minimal stimulation as time (ms) to reach 10% of peak I_{AMPA} from stimulation artefact (wild type: 3.0 ± 0.3 ($n = 12$ cells/ $N = 6$ animals); $K^bD^{b-/-}$: 2.7 ± 0.1 ($n = 23/N = 8$); $K^bD^{b-/-}; NSED^{b+}$: 3.0 ± 0.2 ($n = 17/N = 7$); $K^bD^{b-/-}; NSED^{b-}$: 2.6 ± 0.2 ($n = 19/N = 5$); $P > 0.5$, t -test). **d,** Cumulative probability histogram shows no difference in Max-AMPA between wild type and $K^bD^{b-/-}$. Inset: mean \pm s.e.m. Wild type: 2.6 ± 0.4 nA ($n = 14/N = 6$); $K^bD^{b-/-}$: 2.9 ± 0.4 nA ($n = 22/N = 8$); $P > 0.1$, Mann-Whitney U -test. **e-h,** Presynaptic release probability at $K^bD^{b-/-}$ retinogeniculate synapses is similar to wild type at P20–24. **e, f,** Examples of EPSCs evoked by paired-pulse stimulation of optic tract (20 Hz) recorded in

whole-cell mode in individual LGN neurons from wild type (**e**) versus $K^bD^{b-/-}$ (**f**). **g, h,** Paired-pulse depression (PPD) (%) (EPSC 2/EPSC 1) over varying intervals. **g,** Wild type (open circle) versus $K^bD^{b-/-}$ (filled circle) without cyclothiazide (CTZ), a blocker of AMPA receptor desensitization. Wild type versus $K^bD^{b-/-}$: 1 Hz: 82.3 ± 2.6 ($n = 10$) versus 77.4 ± 3.4 ($n = 8$); 10 Hz: 44.9 ± 2.3 ($n = 9$) versus 42.8 ± 3.9 ($n = 9$); 20 Hz: 37.1 ± 2.1 ($n = 10$) versus 37.3 ± 2.8 ($n = 9$) ($P > 0.1$ for each). **h,** Wild type (open circle) versus $K^bD^{b-/-}$ (filled circle) with CTZ (20 μ M). Wild type versus $K^bD^{b-/-}$: 1 Hz: 79.0 ± 2.4 ($n = 9$) versus 81.8 ± 1.1 ($n = 7$); 10 Hz: 59.2 ± 3.0 ($n = 8$) versus 57.4 ± 3.7 ($n = 7$); 20 Hz: 58.5 ± 2.3 ($n = 8$) versus 56.6 ± 4.1 ($n = 7$) ($P > 0.1$ for each). $N = 4$ for wild type; $N = 3$ for $K^bD^{b-/-}$ for **g, h**. There was no significant difference in PPD between wild type and $K^bD^{b-/-}$, but note significant decrease of PPD +20 μ M CTZ versus 0 μ M CTZ application for both wild type and $K^bD^{b-/-}$ at 10 Hz and 20 Hz ($P < 0.05$). t -test. mean \pm s.e.m. $n =$ cells/ $N =$ animals.



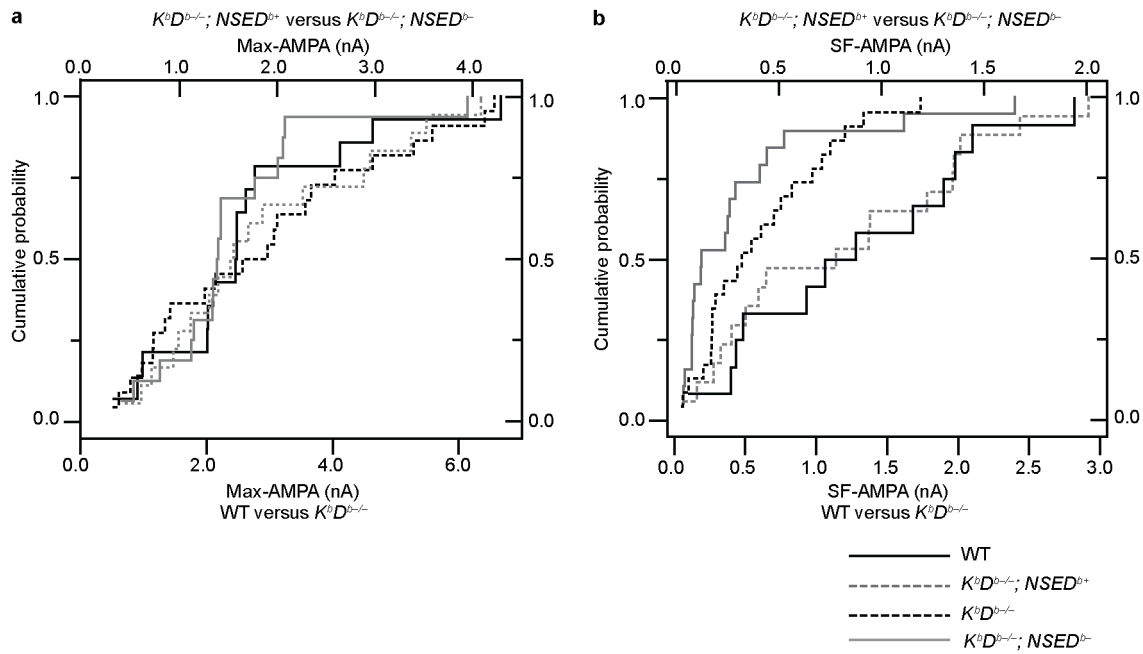
Extended Data Figure 2 | Intact spatio-temporal pattern of retinal waves in $K^bD^{b-/-}$ mice at P5–P12. **a**, Correlation indices as a function of inter-electrode distance for all cell pairs for wild type (grey) and $K^bD^{b-/-}$ (black) at P5–P8. Data points correspond to mean values of medians from individual data sets and error bars represent s.e.m. **b–e**, Summary of temporal firing patterns for retinas isolated from wild-type (open squares) and $K^bD^{b-/-}$ (filled squares) mice at P5–P8 (stage II, cholinergic waves) and P10–P12 (stage III, glutamatergic waves). Open circles correspond to the mean values of individual retinas. **b**, Interburst interval for stage II (in min): wild type: 1.9 ± 0.2 ; $K^bD^{b-/-}$:

1.6 ± 0.2 ; stage III: wild type: 1.0 ± 0.3 ; $K^bD^{b-/-}$: 0.9 ± 0.1 . **c**, Firing rate during burst for stage II (in Hz): wild type: 8.6 ± 1.9 ; $K^bD^{b-/-}$: 8.4 ± 1.7 ; stage III: wild type: 19.7 ± 2.1 ; $K^bD^{b-/-}$: 18.6 ± 2.2 . **d**, Burst duration for stage II (in seconds): wild type: 4.2 ± 0.2 ; $K^bD^{b-/-}$: 3.8 ± 0.1 ; stage III: wild type: 2.6 ± 0.6 ; $K^bD^{b-/-}$: 2.1 ± 0.2 . **e**, Mean firing rate for stage II (in Hz): wild type: 0.5 ± 0.1 ; $K^bD^{b-/-}$: 0.5 ± 0.2 ; stage III: wild type: 1.0 ± 0.1 ; $K^bD^{b-/-}$: 0.8 ± 0.1 , mean \pm s.e.m. ($P > 0.05$ for each, t -test, $N = 6$ animals for each group, except when $N = 5$ for stage III wild type; non-blind experiments).



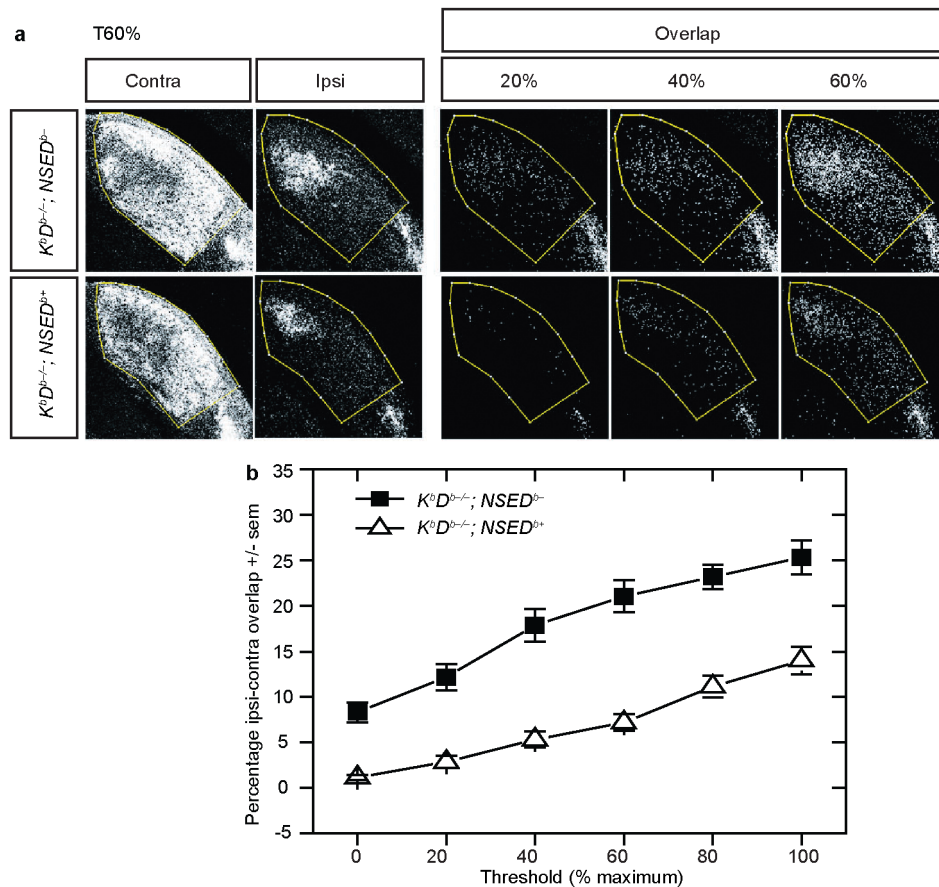
Extended Data Figure 3 | Rescue of H2-D^b expression in brain of $K^bD^{b-/-};NSED^{b+}$ mice. **a**, Diagram of breeding strategy to generate $K^bD^{b-/-};NSED^{b+}$ mice. $K^bD^{b-/-}$ (white indicates absence of H2-D^b) were crossed to $NSED^b$ transgenic mice (black indicates presence of H2-D^b in both body and brain). From F₁ offspring, $H2-K^{b+/-}H2-D^{b+/-};NSED^{b+}$ mice (grey body, black brain) were selected and crossed to $K^bD^{b-/-}$ mice further, generating $K^bD^{b-/-};NSED^{b+}$ (black brain with white body indicates rescue of H2-D^b expression in brain alone) and $K^bD^{b-/-};NSED^{b-/-}$ littermate controls (white body, white brain). **b**, H2-D^b-specific primers. Solid arrows: forward (exon 2) and reverse (exon 3) for ~200 bp spliced mRNA as well as unspliced proRNA (~500 bp) (e; exon, i; intron). Dotted arrows: $NSED^b$ -A (forward, NSE promoter region) and $NSED^b$ -B1 and/or B2 (reverse, exon 2) for genotyping and mRNA detection²⁰. **c**, RT-PCR showing rescue at P10 in thalamus of $K^bD^{b-/-};NSED^{b+}$ mice cDNAs. Wild-type thalamus (WT(T)) shown as positive control; $K^bD^{b-/-}$ thalamus ($K^bD^{b-/-}$ (T)) and spleen ($K^bD^{b-/-}$ (S)) as

negative controls, and various organs from $K^bD^{b-/-};NSED^{b+}$ mice (spleen (S), liver (L), gut (G), thalamus (T), hippocampus (H), cortex (C), retina (R)) were used as templates. **d**, Quantitative PCR comparing relative H2-K^b/H2-D^b gene expression in wild type, $K^bD^{b-/-}$ and $K^bD^{b-/-};NSED^{b+}$ thalami. Left: results show small but highly significant rescue of H2-D^b mRNA expression in $K^bD^{b-/-};NSED^{b+}$ ($K^bD^{b-/-}$: 0.00003 ± 0.00002 ; $K^bD^{b-/-};NSED^{b+}$: 0.0035 ± 0.00044 relative to wild type: 1.0056 ± 0.032 , *** $P = 0.0001$). Each point represents average relative gene expression for one mouse. Right: raw H2-K^b/H2-D^b Ct values for each genotype (wild type: 27.4 ± 0.1 ; $K^bD^{b-/-}$: 44.3 ± 0.4 ; $K^bD^{b-/-};NSED^{b+}$: 35.9 ± 0.3 ; *** $P < 0.001$), one-way ANOVA, mean \pm s.e.m. $N = 5$ animals for wild type, 7 for $K^bD^{b-/-}$, 10 for $K^bD^{b-/-};NSED^{b+}$. **e**, Rescue of H2-D^b protein in $K^bD^{b-/-};NSED^{b+}$ brains at P60. Western blot of immunoprecipitation from whole brain (three pooled brains) lysate from $K^bD^{b-/-}$ and $K^bD^{b-/-};NSED^{b+}$. H2-D^b-specific signal from $K^bD^{b-/-};NSED^{b+}$ appears below IgG band.



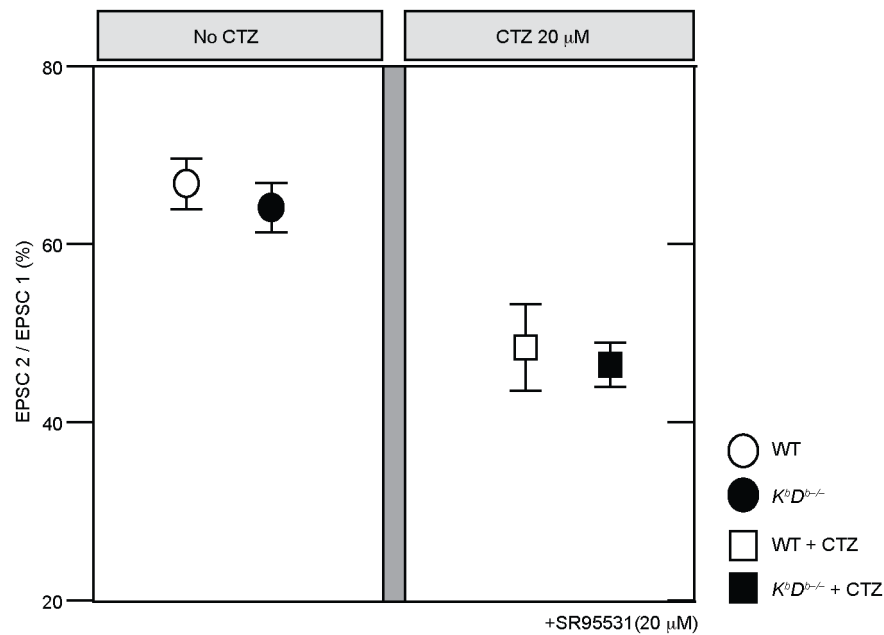
Extended Data Figure 4 | Cumulative probability distribution for SF-AMPA and Max-AMPA recorded at retinogeniculate synapses according to $H2-D^b$ genotype. a, b, Max-AMPA (a) and SF-AMPA (b) showed similar cumulative probability histograms between wild type (black line), $K^bD^{b-/-}; NSED^{b+}$ (dashed grey line), $K^bD^{b-/-}$ (dashed black line) and $K^bD^{b-/-}; NSED^{b-}$ (grey line). Number of experiments are the same as in the

main text, except for Max-AMPA: for $K^bD^{b-/-}; NSED^{b-}$: $n = 18$ cells/ $N = 5$ animals; for $K^bD^{b-/-}; NSED^{b+}$: $n = 16$ / $N = 7$ ($P > 0.05$, Mann–Whitney U -test). Fibre fraction calculated from Max-AMPA and SF-AMPA measurements is similar between wild-type and $K^bD^{b-/-}; NSED^{b+}$ mice (Fig. 2d).



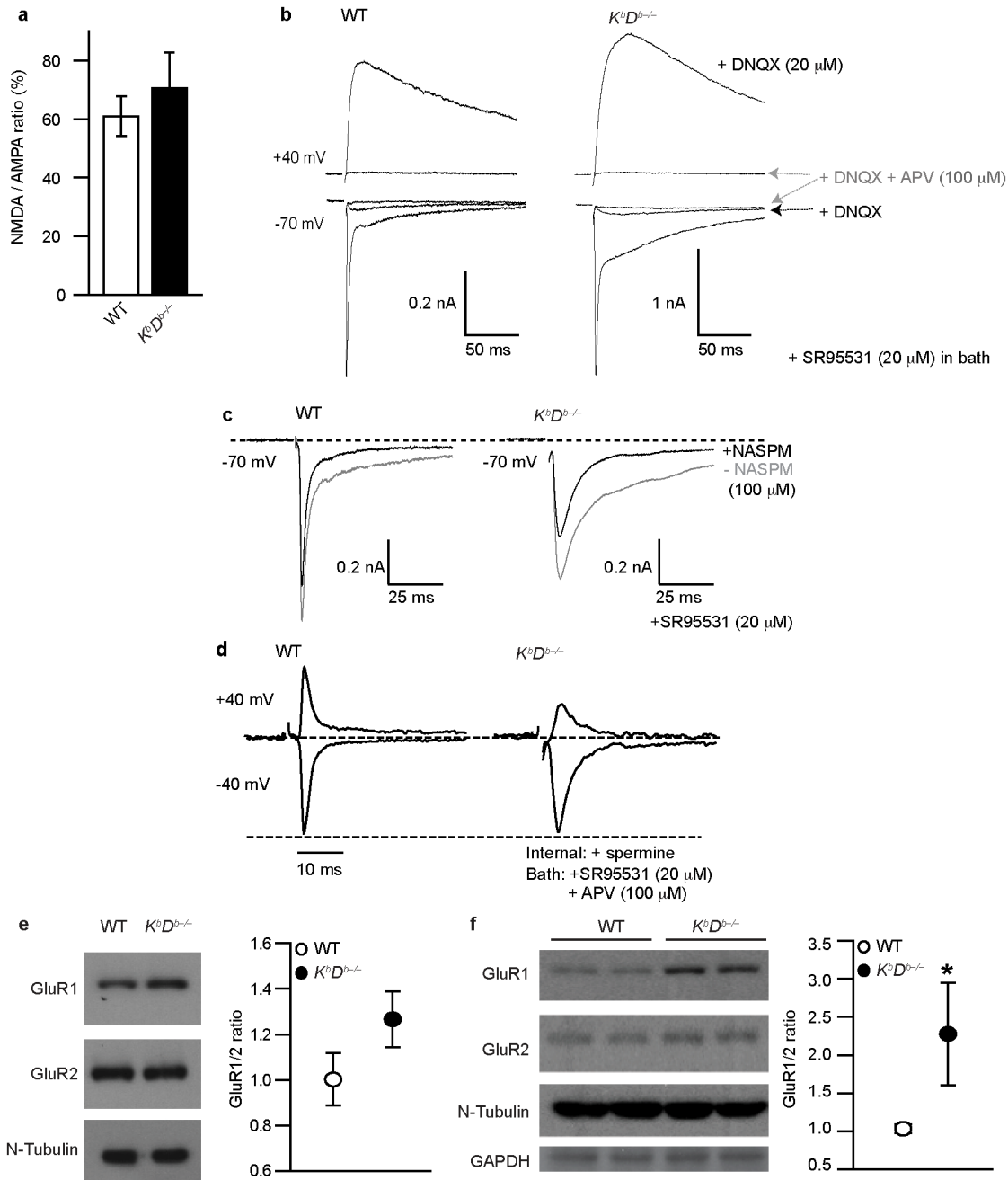
Extended Data Figure 5 | Neuronal H2-D^b expression in $K^bD^{b-/-}$ mice rescues impaired eye-specific axonal segregation at P34. a, Coronal sections of dLGN of $K^bD^{b-/-}; NSED^{b-}$ (top row) and $K^bD^{b-/-}; NSED^{b+}$ (rescue; bottom row) showing pattern of RGC axonal projections from the two eyes after intraocular tracer injections of CTB AF594 (red channel; contralateral eye injected) or AF488 (green channel; ipsilateral eye injected). Left: thresholded fluorescent images of dLGN at 60% maximum signal intensity (see Fig. 2e). Right: overlap of RGC projections (white pixels) from ipsilateral and contralateral eyes displayed for 20%, 40% and 60% maximal threshold

for $K^bD^{b-/-}; NSED^{b-}$ (top) and $K^bD^{b-/-}; NSED^{b+}$ (bottom). Overlap = pixels labelled in both red and green channels. **b,** Mean percentage dLGN area ± s.e.m. pixel overlap for $K^bD^{b-/-}; NSED^{b-}$ (filled squares; $N = 3$) versus $K^bD^{b-/-}; NSED^{b+}$ (open triangles; $N = 4$): 0% threshold: 8.3 ± 1.1 versus 1.2 ± 0.3 ; 20% threshold: 12.2 ± 1.5 versus 2.9 ± 0.7 ; 40% threshold: 17.9 ± 1.8 versus 5.3 ± 0.9 ; 60% threshold: 21.1 ± 1.8 versus 7.2 ± 0.9 ; 80% threshold: 23.2 ± 1.3 versus 11.1 ± 1.2 ; 100%: 25.3 ± 1.9 versus 14.0 ± 1.5 ($P < 0.05$, two-way ANOVA). See Methods and ref. 47.



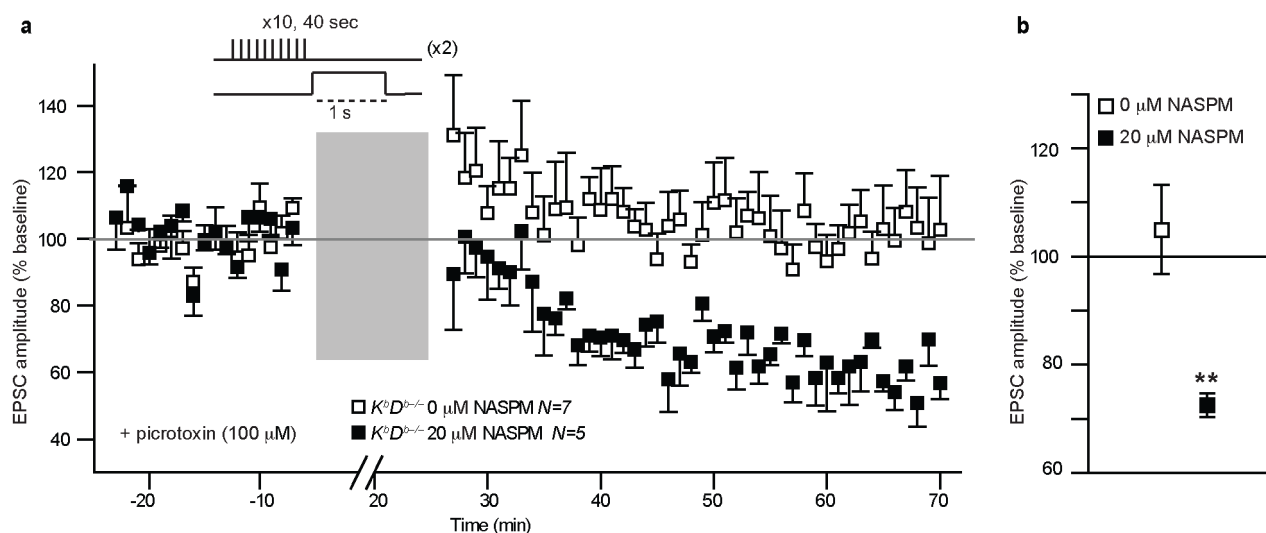
Extended Data Figure 6 | Intact release probability at $K^bD^{b-/-}$ retinogeniculate synapses before eye opening. Paired-pulse stimulation was delivered to the optic tract at 10 Hz, similar to the natural firing frequency of RGCs (Extended Data Fig. 2c), and whole-cell recordings were made from LGN neurons in slices aged between P8–13. Paired-pulse stimulation resulted in synaptic depression, represented as EPSC 2 divided by EPSC 1 (%). In 0 μM CTZ (left panel): wild type: 67.0 ± 2.9 ($n = 11/N = 4$); $K^bD^{b-/-}$: 64.2 ± 2.8 ($n = 7/N = 2$). In 20 μM CTZ (right panel): wild type: 48.6 ± 4.9 ($n = 8/N = 3$);

$K^bD^{b-/-}$: 46.6 ± 2.5 ($n = 7/N = 2$) ($P > 0.1$ for each, t -test), mean \pm s.e.m. 20 mM BAPTA containing Cs^+ -internal solution was used for this experiment due to prolonged kinetics of EPSCs in $K^bD^{b-/-}$. The identical paired-pulse ratios between wild type and $K^bD^{b-/-}$ are consistent with the conclusion that presynaptic release probability is intact at P8–13 retinogeniculate synapses in $K^bD^{b-/-}$ mice. (See also Extended Data Fig. 1e–h for similar conclusion at P20–24, after synapse elimination is largely complete.)



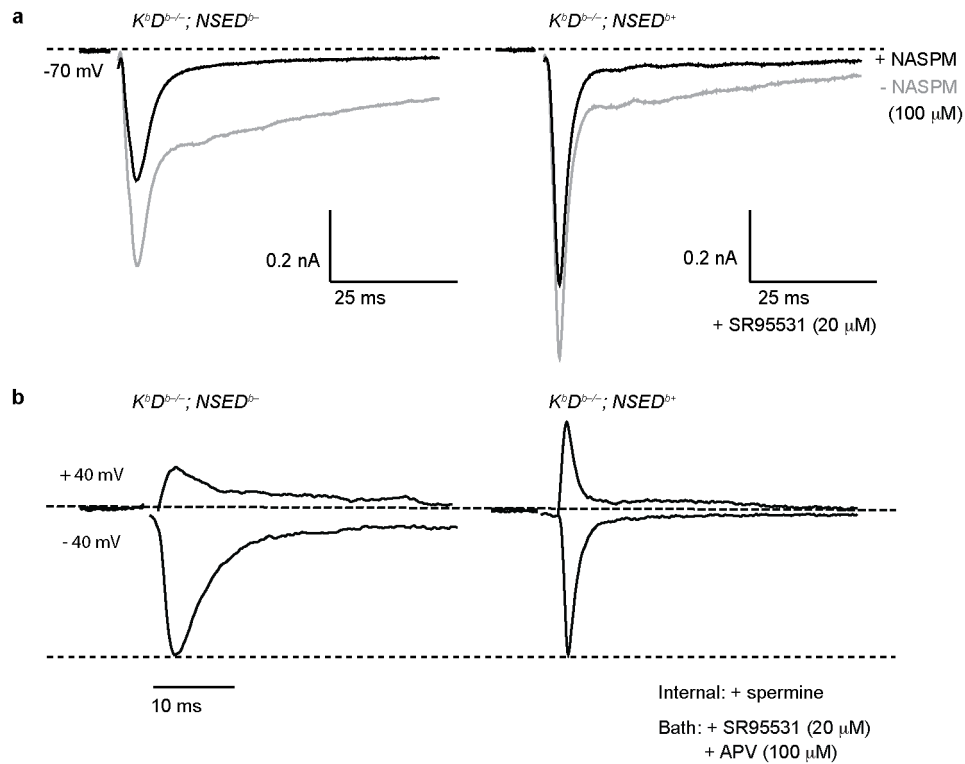
Extended Data Figure 7 | Normal NMDA/AMPA ratio but increased Ca^{2+} -permeable AMPA receptors at retinogeniculate synapses in $K^bD^{b-/-}$ mice. **a, b**, NMDA/AMPA ratio is unchanged in $K^bD^{b-/-}$ mice. **a**, NMDA/AMPA ratio (%): peak I_{AMPA} measured at -70 mV ($+20$ μ M SR95531) versus peak I_{NMDA} at $+40$ mV ($+20$ μ M SR95531 + 20 μ M DNQX): wild type: 61 ± 6.8 ($n = 10/N = 4$); $K^bD^{b-/-}$: 70.6 ± 12.1 ($n = 7/N = 3$) ($P > 0.1$, t -test) mean \pm s.e.m. **b**, Example recordings from individual neurons for wild type (left) and $K^bD^{b-/-}$ (right). APV (100 μ M) was added at the end of each experiment to confirm NMDA-mediated synaptic currents. D600 in pipette. **c**, Example showing effect of NASPM (100 μ M bath) on I_{AMPA} : note significant blockade of I_{AMPA} in $K^bD^{b-/-}$. Grey line, before NASPM; black line, after

NASPM (5 traces averaged for single cell). SR95531 (20 μ M) added to bath for **a–c**. **d**, Examples for I_{AMPA} normalized to EPSC amplitude at -40 mV. Note reduction in EPSC amplitude at $+40$ mV in $K^bD^{b-/-}$ but not wild type. 100 μ M APV + 20 μ M SR95531 in bath. Spermine (100 μ M) and D600 (100 μ M) in pipette. Ages: P8–13. Experimenter was aware of genotype due to obvious differences in time course of EPSCs and effects of NASPM. **e**, Example western blot (left) and GluR1/GluR2 ratio (right) of P22 thalamus; wild type: 1.0 ± 0.1 ($N = 12$); $K^bD^{b-/-}$: 1.3 ± 0.1 ($N = 13$) ($P = 0.07$). **f**, Example western blot (left) and GluR1/GluR2 ratio (right) of cultured cortical neurons; wild type: 1.0 ± 0.1 ($N = 4$); $K^bD^{b-/-}$: 2.3 ± 0.7 ($N = 4$) ($*P = 0.03$). Mann–Whitney U -test for **e, f**, $n = \text{cells}/N = \text{animals}$.



Extended Data Figure 8 | NASPM-dependent rescue of LTD in $K^bD^{b-/-}$ LGN at P8–13. **a**, Summary of all 1,100 ms latency experiments: EPSC peak amplitude (% change from baseline) versus time ($n = 7/N = 7$ for 0 μM NASPM; $n = 5/N = 5$ for 20 μM NASPM). Grey bar = induction period

(Methods). 1 min data binning. **b**, Average of per cent change (mean \pm s.e.m.); $K^bD^{b-/-}$ 0 μM NASPM: 105 \pm 8.2 ($N=7$); $K^bD^{b-/-}$ +20 μM NASPM: 72.5 \pm 2.2 ($N=5$). ** $P < 0.01$, t -test, $n = \text{cells}/N = \text{animals}$.



Extended Data Figure 9 | Neuronal H2-D^b expression decreases Ca²⁺ permeability of AMPA receptors at retinogeniculate synapses in $K^bD^{b-/-}; NSED^{b+/+}$ mice. **a**, NASPM blockade of I_{AMPA} is significantly reduced in $K^bD^{b-/-}; NSED^{b+/+}$ mice. Example trace of NASPM effect on I_{AMPA} recorded from $K^bD^{b-/-}; NSED^{b-/-}$ (left) or $K^bD^{b-/-}; NSED^{b+/+}$ (right) individual LGN neuron. Grey line, before NASPM; black line, after NASPM application (5 traces averaged for single cell); 20 μ M SR95531 in bath. **b**, Internal

spermine-dependent block of I_{AMPA} at positive membrane potentials is rescued in $K^bD^{b-/-}; NSED^{b+/+}$ LGN neurons. Example recordings for I_{AMPA} normalized to EPSC amplitude at -40 mV from individual neurons. Note reduction in EPSC amplitude at +40 mV in $K^bD^{b-/-}; NSED^{b-/-}$, but restored to wild-type level in $K^bD^{b-/-}; NSED^{b+/+}$. 100 μ M APV + 20 μ M SR95531 in bath. Spermine (100 μ M) and D600 (100 μ M) in internal solution.

Circular polarization in the optical afterglow of GRB 121024A

K. Wiersema¹, S. Covino², K. Toma^{3,4,5}, A. J. van der Horst⁶, K. Varela⁷, M. Min⁶, J. Greiner⁷, R. L. C. Starling¹, N. R. Tanvir¹, R. A. M. J. Wijers⁶, S. Campana², P. A. Curran⁸, Y. Fan⁹, J. P. U. Fynbo¹⁰, J. Gorosabel^{11,12,13}, A. Gomboc¹⁴, D. Götz¹⁵, J. Hjorth¹⁰, Z. P. Jin⁹, S. Kobayashi¹⁶, C. Kouveliotou¹⁷, C. Mundell¹⁶, P. T. O'Brien¹, E. Pian^{18,19}, A. Rowlinson⁶, D. M. Russell^{20,21,22}, R. Salvaterra²³, S. di Serego Alighieri²⁴, G. Tagliaferri², S. D. Vergani², J. Elliott⁷, C. Fariña²⁵, O. E. Hartoog⁶, R. Karjalainen²⁵, S. Klose²⁶, F. Knust⁷, A. J. Levan²⁷, P. Schady⁷, V. Sudilovsky⁷ & R. Willingale¹

Gamma-ray bursts (GRBs) are most probably powered by collimated relativistic outflows (jets) from accreting black holes at cosmological distances. Bright afterglows are produced when the outflow collides with the ambient medium. Afterglow polarization directly probes the magnetic properties of the jet when measured minutes after the burst, and it probes the geometric properties of the jet and the ambient medium when measured hours to days after the burst^{1–5}. High values of optical polarization detected minutes after the burst of GRB 120308A indicate the presence of large-scale ordered magnetic fields originating from the central engine⁵ (the power source of the GRB). Theoretical models predict low degrees of linear polarization and no circular polarization at late times^{6–8}, when the energy in the original ejecta is quickly transferred to the ambient medium and propagates farther into the medium as a blast wave. Here we report the detection of circularly polarized light in the afterglow of GRB 121024A, measured 0.15 days after the burst. We show that the circular polarization is intrinsic to the afterglow and unlikely to be produced by dust scattering or plasma propagation effects. A possible explanation is to invoke anisotropic (rather than the commonly assumed isotropic) electron pitch-angle distributions, and we suggest that new models are required to produce the complex microphysics of realistic shocks in relativistic jets^{9–11}.

Magnetic fields play a crucial role in the physics of relativistic jets—for example, in their formation, acceleration and collimation^{1,12,13}. On smaller spatial scales, there is a strong connection between particle acceleration and magnetic field generation in the collisionless relativistic shocks that create GRB afterglows^{9–11}. Our understanding of magnetic field properties in GRBs and their afterglows has improved rapidly through recent observational successes, such as time-resolved linear polarimetry of prompt γ -ray emission^{14,15} (just seconds after the burst) and the observed transition of reverse shock emission (high levels of optical linear polarization, seen just minutes after the burst) to the early forward shock afterglow emission (lower levels of linear polarization)⁵. These observations have given support to models predicting large-scale ordered fields in the GRB ejecta⁵. On the other hand, late-time polarimetry (hours to days after the burst) offers the advantage that at these times the optical emission probably originates from a single emission process, namely synchrotron

emission from the forward shock (that is, emission from the shocked ambient medium), in which ordered fields are much less likely to be present⁵. This then allows measurements of the geometry of the jet through monitoring of the late-time polarization angle^{2–4,16,17}, but crucially also offers a simple test for afterglow microphysics via circular polarimetry—detailed models have been developed, but observational attempts have largely focused on radio reverse shocks, producing only upper limits¹⁸, and a single optical non-detection¹⁷.

GRB 121024A was detected by the Burst Alert Telescope (BAT) on board the Swift satellite at 02:56:12 UT on 24 October 2012¹⁹; a redshift of $z = 2.298$ was determined shortly afterwards (Methods). We obtained imaging polarimetry observations with the FOcal Reducer and low dispersion Spectrograph (FORIS2) on the Very Large Telescope (VLT), using a Wollaston prism and quarter- and half-wavelength plates (Extended Data Figs 1 and 2). Observations with the ESO R_{special} filter commenced at 2.57 h after the burst, when the afterglow had an R-band brightness of $R \approx 19.8$ mag. After two sets of linear polarimetry (giving the percentage of linear polarization, P_{lin}), we obtained four consecutive measurements of the percentage of circular polarization (P_{cir}), followed by a further 9 measurements of P_{lin} (of which 4 were on the second night). Data reduction and calibration follow standard procedures¹⁷ (Methods). Simultaneously with the polarimetry, we monitored the afterglow optical light curve with the Gamma-Ray Burst Optical/Near-Infrared Detector (GROND) instrument (Methods). In the following we will use a notation where the flux density F depends on frequency ν and time t as $F \propto t^{-\alpha} \nu^{-\beta}$, with temporal decay index α and spectral energy index β .

The X-ray light curve obtained by Swift (retrieved from the online Swift/XRT GRB light curve repository²⁰) is well described by three power-law segments, where the first break occurs at $t_{\text{break},1} = 619_{-348}^{+199}$ s and the second break at $t_{\text{break},2} = 3.4_{-2.2}^{+1.5} \times 10^4$ s (errors at 90% confidence level). In the GROND light curve, we find evidence of a break at a time consistent with this last X-ray break (Extended Data Fig. 3). A combined fit to the X-ray and GROND data, using a smoothly broken power law and a host galaxy component, gives a best-fitting break time $t_{\text{break},2} = 3.72 \pm 0.07 \times 10^4$ s (Methods), pre-break light-curve decay indices $\alpha_{\text{pre,opt}} = 0.93 \pm 0.02$ and $\alpha_{\text{pre,X-ray}} = 0.96 \pm 0.11$, and post-break light-curve decay indices $\alpha_{\text{post,opt}} = 1.25 \pm 0.04$ and $\alpha_{\text{post,X-ray}} = 1.67 \pm 0.10$ (uncertainties

¹Department of Physics and Astronomy, University of Leicester, Leicester LE1 7RH, UK. ²INAF/Brera Astronomical Observatory, via Biancamano 46, I-23807 Merate (LC), Italy. ³Department of Earth and Space Science, Osaka University, Toyonaka 560-0043, Japan. ⁴Astronomical Institute, Tohoku University, Sendai 980-8578, Japan. ⁵Frontier Research Institute for Interdisciplinary Sciences, Tohoku University, Sendai 980-8578, Japan. ⁶Astronomical Institute 'Anton Pannekoek', University of Amsterdam, PO Box 94248, 1090 SJ Amsterdam, The Netherlands. ⁷Max-Planck-Institut für extraterrestrische Physik, Giessenbachstrasse 1, D-85748 Garching, Germany. ⁸International Centre for Radio Astronomy Research, Curtin University, GPO Box U1987, Perth, Western Australia 6845, Australia. ⁹Key Laboratory of Dark Matter and Space Astronomy, Purple Mountain Observatory, Chinese Academy of Science, Nanjing 210008, China. ¹⁰Dark Cosmology Centre, Niels Bohr Institute, University of Copenhagen, Juliane Maries Vej 30, DK 2100 Copenhagen, Denmark. ¹¹Instituto de Astrofísica de Andalucía (IAA-CSIC), Glorieta de la Astronomía s/n, E-18008 Granada, Spain. ¹²Unidad Asociada Grupo Ciencia Planetarias UPV/EHU-IAA-CSIC, Departamento de Física Aplicada I, ETS Ingeniería, Universidad del País Vasco UPV/EHU, Alameda de Urquijo s/n, E-48013 Bilbao, Spain. ¹³Kerbasque, Basque Foundation for Science, Alameda de Urquijo 36-5, E-48008 Bilbao, Spain. ¹⁴Faculty of Mathematics and Physics, University of Ljubljana, Jadranska 19, 1000 Ljubljana, Slovenia. ¹⁵AIM (UMR 7158 CEA/DSM-CNRS-Université Paris Diderot) Irfu/Service d'Astrophysique, Saclay, F-91191 Gif-sur-Yvette Cedex, France. ¹⁶Astrophysics Research Institute, Liverpool John Moores University, Liverpool Science Park, IC2 Building, 146 Brownlow Hill, Liverpool L3 5RF, UK. ¹⁷Space Science Office, ZP12, NASA/Marshall Space Flight Center, Huntsville, Alabama 35812, USA. ¹⁸Scuola Normale Superiore, 7, I-56126 Pisa, Italy. ¹⁹INAF/IASF Bologna, via Gobetti 101, I-40129 Bologna, Italy. ²⁰Instituto de Astrofísica de Canarias (IAC), E-38200 La Laguna, Tenerife, Spain. ²¹Departamento de Astrofísica, Universidad de La Laguna, E-38206 La Laguna, Tenerife, Spain. ²²New York University Abu Dhabi, PO Box 129188, Abu Dhabi, United Arab Emirates. ²³INAF/IASF Milano, via E. Bassini 15, 20133 Milano, Italy. ²⁴INAF-Osservatorio Astrofisico di Arcetri, Largo E. Fermi 5, I-50125 Firenze, Italy. ²⁵Isaac Newton Group of Telescopes, Apartado de Correos 321, E-38700 Santa Cruz de la Palma, Canary Islands, Spain. ²⁶Thüringer Landessternwarte Tautenburg, Sternwarte 5, 07778 Tautenburg, Germany. ²⁷Department of Physics, University of Warwick, Coventry CV4 7AL, UK.

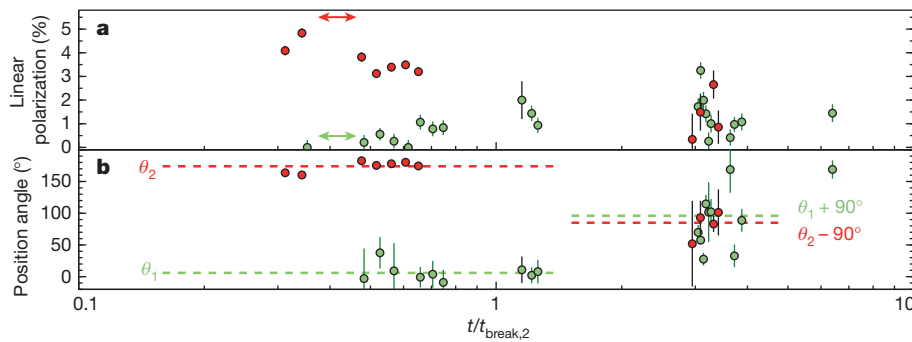


Figure 1 | Linear polarization of the afterglow of GRB 121024A. **a**, The percentage of linear polarization as a function of time; **b**, the polarization angle. Red points are for GRB 121024A, green points are for the only other Swift GRB afterglow with extensive polarimetry, GRB 091018¹⁷ (error bars are 1σ). The horizontal axis marks the time since trigger, normalized by the jet break time

($t_{\text{break},2} = 3.72 \pm 0.07 \times 10^4$ s). θ_1 and θ_2 show the average angles before the jet break; $\theta_{1,2} + 90^\circ$ demonstrates the 90° angle change predicted for jet breaks of uniform jets. Horizontal double-headed arrows show the timespan over which the circular polarimetry was obtained.

are 1σ). The occurrence of a light-curve break simultaneously in X-ray and optical wavelengths is suggestive of a jet break origin: a late-time, usually achromatic, light-curve break, generally associated with the finite opening angle of the jet that becomes apparent when the jet Lorentz factor decreases below a critical value. The X-ray+GROND spectral energy distribution is best fitted with a single power law with $\beta = 0.88 \pm 0.01$ and an optical extinction $A_V = 0.22 \pm 0.02$ mag (Methods; Extended Data Fig. 4). The pre-break temporal and spectral indices agree with the standard fireball closure relations²¹ in the situation where the synchrotron cooling frequency ν_c and peak frequency ν_m are below both optical and X-ray frequencies²¹.

Figure 1 shows the observed P_{lin} behaviour. Initially the source starts out $\sim 5\%$ polarized (much lower than the $\sim 70\%$ expected for a perfectly ordered magnetic field); this polarization subsequently decays to lower levels, while the polarization angle is remarkably constant. This shows that the magnetic field directions are largely random, that is, the coherence scales of the field in the blast wave are small, but their directions are confined to the plane of the shock (the detected polarization is attributed to a somewhat off-axis viewing angle). In the second night of data, the polarization angle is markedly different, consistent (within errors) with a change of exactly 90° (Fig. 1). We consider this an unambiguous detection of the 90° angle change predicted to occur around the jet break time of a homogeneous jet that is not spreading sideways⁴. The exact time at which the angle change occurs is dictated by viewing angle⁴. The observed angle change shows that any ordered magnetic fields in the forward shock are weak, if present²².

We acquired the P_{cir} measurements shown in Fig. 2 between the second and third P_{lin} datapoints. Under the assumption that during this interval no variability in P_{cir} is expected (the time covered is small compared to the time after the burst), we combine the 4 measurements together, and measure $P_{\text{cir}} = 0.61\% \pm 0.13\%$ (see Methods). We estimate the percentage of linear polarization during the circular polarimetry interval to be $P_{\text{lin}} \approx 4\%$: the afterglow shows a ratio $P_{\text{cir}}/P_{\text{lin}} \approx 0.15$, several orders of magnitude above basic model predictions ($\sim 10^{-4}$ at optical wavelengths^{6–8}; see Methods) and other measured relativistic jet sources²³ (Fig. 3).

A high level of circular polarization can be intrinsic to the source (for example, the P_{cir} of the synchrotron emission from the source) or have its origin in propagation effects within the source (for example, Faraday conversion $P_{\text{lin}} \rightarrow P_{\text{cir}}$, which is effective in a hot, relativistic medium) or by dust scattering effects along the line of sight. Plasma propagation effects within the source have been shown to be strong at long wavelengths (close to the synchrotron self-absorption frequency, ν_a), but negligible at optical wavelengths^{6–8}; it is very unlikely that these effects play a role here.

The influence of dust is limited to the host galaxy, as the Galactic extinction towards GRB 121024A is very small, $E(B - V) = 0.10$ (Methods). Dust affects optical P_{cir} through four possible routes: multiple scattering in an optically thick medium of dust grains; dichroic scattering by

(somewhat aligned) non-spherical dust grains; dichroic extinction of linearly polarized radiation by (somewhat aligned) non-spherical dust grains; and the scattering of linearly polarized radiation by randomly oriented dust particles (if the polarization is not in, or perpendicular to, the scattering plane). The P_{cir} expected to be caused by the last three effects depends strongly on the degree of alignment, source inclination and the total amount of dust involved, for which line-of-sight extinction may be a proxy^{24,25}. The weak line-of-sight host galaxy extinction, $A_V = 0.22 \pm 0.02$ mag (Methods), argues against multiple-scattering effects²⁵. To rule out the other possibilities, we simulated the efficiency of conversion of $P_{\text{lin}} \rightarrow P_{\text{cir}}$ through single dust scattering (that is, where each photon has been scattered at most once by a dust particle), finding maximal conversion values at high and low scattering angle regimes of $\sim 10\%$ at $\sim 100^\circ$, and $\sim 8\%$ at $\sim 20^\circ$. Models based on partially aligned dust grains have lower efficiency. If a large part of P_{cir} is in fact due to dust, we expect a large fraction, if not all, of the P_{lin} to be caused by dust scattering too. The linear polarization curve (Fig. 1) shows no signs of this: we would not expect to see a constant angle before the jet break (at $t_{\text{break},2}$), nor for P_{lin} to reach near zero at any point, nor a clear 90° angle change over the jet break²⁶. In addition, light scattered at large scattering angles has not had time to reach us yet (dust very close to the GRB gets destroyed by the GRB prompt emission). After eliminating plasma propagation effects and dust scattering, we therefore conclude that the measured P_{cir} is likely to be largely or fully intrinsic to the afterglow. It is of some interest to compare this with the only other afterglow with deep measurements,

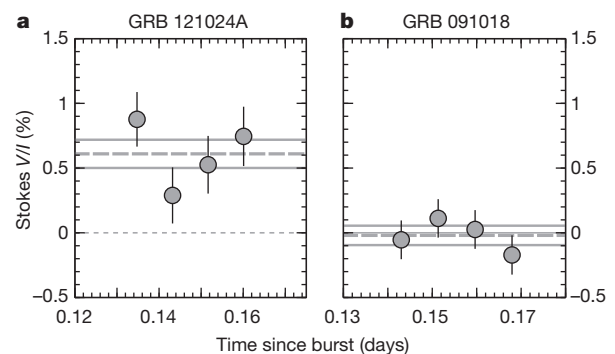


Figure 2 | Optical circular polarization measurements of the afterglow of GRB 121024A. **a**, Optical circular polarimetry of the afterglow of GRB 121024A; **b**, optical circular polarimetry for GRB 091018¹⁷ (error bars are 1σ). The horizontal axis shows the time since burst in the observer frame, and the vertical axis shows the circular polarization expressed in Stokes parameters V and I (see Methods) as V/I , in per cent. The short-dashed line indicates $V/I = 0$, the dashed line shows the measured V/I from the combined data points, and 1σ uncertainty values around the best combined data value are shown with solid lines.

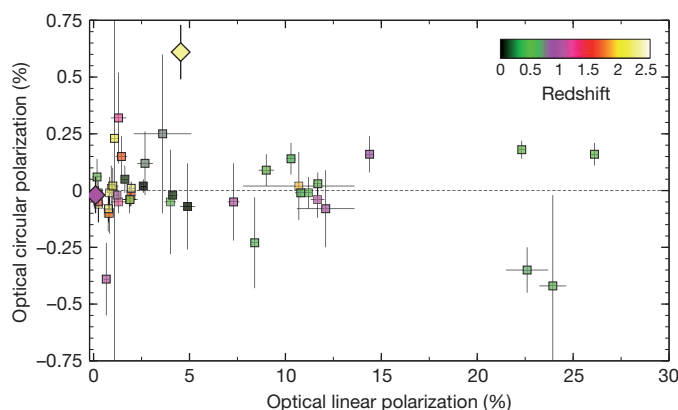


Figure 3 | Optical polarimetry of quasars and GRB afterglows compared. Polarization measurements for GRBs 121024A and 091018 (diamond symbols) are shown with optical polarimetry of quasars²³ (squares); error bars are 1σ . The data points are colour-coded for redshift. The few quasars with $>3\sigma$ detection of optical circular polarization have high values ($\geq 20\%$) of linear polarization. GRB 121024A has the highest detected level of optical circular polarization, and at $z = 2.298$ is the highest-redshift object with a detection, showing the potential for GRBs as probes of possible cosmological propagation effects. We note that the range of redshifts means that different rest-frame wavelengths are being compared.

that of GRB 091018, which showed limits¹⁷ $P_{\text{cir}} < 0.15\%$ (2σ) and $P_{\text{cir}}/P_{\text{lin}} < 1$ (Figs 1, 2 and 3).

The origin of the optical circular polarization in the afterglow of GRB 121024A is puzzling. We expect intrinsic polarization from the external shock synchrotron emission of $P_{\text{cir}} \approx \gamma_e^{-1}$, where γ_e is the random Lorentz factor of the electrons emitting the optical radiation—under the assumption of isotropic electron pitch-angle distribution and perfectly ordered magnetic field⁸ (Methods). In this situation, the observed P_{cir} would imply extremely low values of γ_e . Furthermore, because P_{cir} and P_{lin} are both expected to be reduced from the value for a perfectly ordered field by the field randomness in the same way, the expected $P_{\text{cir}}/P_{\text{lin}}$ ratio is the same as for a perfectly ordered field, again scaling with γ_e^{-1} in the case of an isotropic electron distribution⁸. Therefore, the observed polarimetric behaviour poses a challenge to the long-standing assumption of isotropic electron pitch-angle distributions in the GRB forward shock afterglow. Pitch-angle anisotropy has been postulated before as a possible explanation of GRB phenomena as varied as steep decay phases in X-ray light curves²⁷ and spectra of prompt emission²⁸, but evidence for this has been lacking. Further hints of a more complicated structure of emission and acceleration regions come from observations as varied as high-energy emission in GRBs²⁹ and fast variability of high-energy emission in quasars³⁰. The circular polarization of afterglows from GRBs as well as quasars (Fig. 3) offers a new line of evidence required to guide theoretical studies^{9,10}.

METHODS SUMMARY

Both linear and circular polarimetry of the afterglow of GRB 121024A were performed using the FORS2 instrument on the VLT, using the R_{special} filter in imaging polarimetry mode, using a Wollaston prism and half- and quarter-wavelength plates. We used four plate angles for linear polarimetry and two for circular polarimetry, in order to be able to use beam-switching to reduce systematic errors, and used aperture photometry to measure source fluxes. Measurements from polarimetric sequences taken on the second night after the burst were combined to increase signal-to-noise ratios. We used field stars to measure the linear polarization induced by Galactic dust, fitting the distribution in Stokes Q and U with a two-dimensional Gaussian function. We monitored the optical afterglow brightness in seven photometric filters using the GROND instrument, we fitted the X-ray data from the Swift satellite together with the GROND data with absorbed power laws to establish the presence of a late-time break in the light curve, and we fitted the X-ray/optical spectral energy distribution of the afterglow. Finally, we demonstrated how the observed high level of circular polarization contradicts theoretical estimates, and how

anisotropy in the electron pitch-angle distribution may explain the observed ratio of optical circular to linear polarization.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 21 November 2013; accepted 4 March 2014.

Published online 30 April 2014.

- Piran, T. Magnetic fields in gamma-ray bursts: a short overview. *AIP Conf. Proc.* **784**, 164–174 (2005).
- Sari, R. Linear polarization and proper motion in the afterglow of beamed gamma-ray bursts. *Astrophys. J.* **524**, L43–L46 (1999).
- Ghisellini, G. & Lazzati, D. Polarization light curves and position angle variation of beamed gamma-ray bursts. *Mon. Not. R. Astron. Soc.* **309**, L7–L11 (1999).
- Rossi, E. M., Lazzati, D., Salmonson, J. D. & Ghisellini, G. The polarization of afterglow emission reveals gamma-ray bursts jet structure. *Mon. Not. R. Astron. Soc.* **354**, 86–100 (2004).
- Mundell, C. G. *et al.* Highly polarized light from stable ordered magnetic fields in GRB 120308A. *Nature* **504**, 119–121 (2013).
- Matsumiya, M. & Ioka, K. Circular polarization from gamma-ray burst afterglows. *Astrophys. J.* **595**, L25–L28 (2003).
- Sagiv, A., Waxman, E. & Loeb, A. Probing the magnetic field structure in gamma-ray bursts through dispersive plasma effects on the afterglow polarization. *Astrophys. J.* **615**, 366–377 (2004).
- Toma, K., Ioka, K. & Nakamura, T. Probing the efficiency of electron-proton coupling in relativistic collisionless shocks through the radio polarimetry of gamma-ray burst afterglows. *Astrophys. J.* **673**, L123–L126 (2008).
- Spitkovsky, A. Particle acceleration in relativistic collisionless shocks: Fermi process at last? *Astrophys. J.* **682**, L5–L8 (2008).
- Spitkovsky, A. On the structure of relativistic collisionless shocks in electron-ion plasmas. *Astrophys. J.* **673**, L39–L42 (2008).
- Heddel, C. B. & Nishikawa, K.-I. The influence of an ambient magnetic field on relativistic collisionless plasma shocks. *Astrophys. J.* **623**, L89–L92 (2005).
- Pudritz, R. E., Hardcastle, M. J. & Gabuzda, D. C. Magnetic fields in astrophysical jets: from launch to termination. *Space Sci. Rev.* **169**, 27–72 (2012).
- Lytikov, M. Magnetocentrifugal launching of jets from discs around Kerr black holes. *Mon. Not. R. Astron. Soc.* **396**, 1545–1552 (2009).
- Götz, D., Laurent, P., Lebrun, F., Daigne, F. & Bosnjak, Z. Variable polarization measured in the prompt emission of GRB 041219A using IBIS on board INTEGRAL. *Astrophys. J.* **695**, L208–L212 (2009).
- Yonetoku, D. *et al.* Magnetic structures in gamma-ray burst jets probed by gamma-ray polarization. *Astrophys. J.* **758**, L1–L6 (2012).
- Greiner, J. *et al.* Evolution of the polarization of the optical afterglow of the γ -ray burst GRB 030329. *Nature* **426**, 157–159 (2003).
- Wiersema, K. *et al.* Detailed optical and near-infrared polarimetry, spectroscopy and broadband photometry of the afterglow of GRB 091018: polarization evolution. *Mon. Not. R. Astron. Soc.* **426**, 2–22 (2012).
- Granot, J. & Taylor, G. B. Radio flares and the magnetic field structure in gamma-ray burst outflows. *Astrophys. J.* **625**, 263–270 (2005).
- Pagani, C. *et al.* GRB 121024A: Swift detection of a burst with an optical counterpart. *GCN Circ.* **13886** (2012).
- Evans, P. A. *et al.* Methods and results of an automatic analysis of a complete sample of Swift-XRT observations of GRBs. *Mon. Not. R. Astron. Soc.* **397**, 1177–1201 (2009).
- Sari, R., Piran, T. & Narayan, R. Spectra and lightcurves of gamma-ray burst afterglows. *Astrophys. J.* **497**, L17–L20 (1998).
- Granot, J. & Königl, A. Linear polarisation in gamma-ray bursts: the case for an ordered magnetic field. *Astrophys. J.* **594**, L83–L87 (2003).
- Hutsemekers, D., Borguet, B., Sluse, D., Cabanac, R. & Lamy, H. Optical circular polarization in quasars. *Astron. Astrophys.* **520**, L7 (2010).
- Whitney, B. A. & Wolff, M. J. Scattering and absorption by aligned grains in circumstellar environments. *Astrophys. J.* **574**, 205–231 (2002).
- Fukue, T. *et al.* Near-infrared circular polarimetry and correlation diagrams in the Orion Becklin-Neugebauer/Kleinman-Low region: contribution of dichroic extinction. *Astrophys. J.* **692**, L88–L91 (2009).
- Lazzati, D. *et al.* Intrinsic and dust-induced polarization in gamma-ray burst afterglows: the case of GRB 021004. *Astron. Astrophys.* **410**, 823–831 (2003).
- Beloborodov, A. M., Daigne, F., Mochkovitch, R. & Uhm, Z. L. Is gamma-ray burst afterglow emission intrinsically anisotropic? *Mon. Not. R. Astron. Soc.* **410**, 2422–2427 (2011).
- Lloyd-Ronning, N. M. & Petrosian, V. Interpreting the behavior of time-resolved gamma-ray burst spectra. *Astrophys. J.* **565**, 182–194 (2002).
- Kouveliotou, C. *et al.* NuSTAR observations of GRB 130427A establish a single component synchrotron afterglow origin for the late optical to multi-GeV emission. *Astrophys. J.* **779**, L1 (2013).
- Ghisellini, G., Tavecchio, F., Bodo, G. & Celotti, A. TeV variability in blazars: how fast can it be? *Mon. Not. R. Astron. Soc.* **393**, L16–L20 (2009).

Acknowledgements This work is based on observations made with ESO telescopes at the Paranal Observatory under programme 090.D-0789. We thank all ING staff for their support of ACAM ToO observations. K.W. thanks J. Hinton for discussions. K.W. was

supported by STFC. K.T. was supported by a JSPS Research Fellowship for Young Scientists no. 231446. A.J.v.d.H., R.A.M.J.W. and A.R. were supported by the European Research Council via Advanced Investigator grant no. 247295. R.L.C.S. was supported by a Royal Society Fellowship. Y.F. was supported by the 973 Programme of China, under grant 2013CB837000. D.M.R. was supported by a Marie Curie Intra European Fellowship within the 7th European Community Framework Programme under contract no. IEF 274805. This work was supported by the Australian Research Council (grant DP120102393). The William Herschel telescope and its override programme are operated on the island of La Palma by the Isaac Newton Group in the Spanish Observatorio del Roque de los Muchachos of the Instituto de Astrofísica de Canarias. This work made use of data supplied by the UK Swift Science Data Centre at the University of Leicester, funded by the UK Space Agency.

Author Contributions K.W. and S.C. jointly led the VLT observing time proposals and defined the observing strategy. K.W. acquired, reduced and analysed the VLT data and took primary responsibility for writing the text of the paper; S.C. performed an independent data analysis. K.T., A.J.v.d.H. and M.M. provided the theoretical interpretation of the observations. K.V. and J.G. analysed the GROND data. O.E.H. led the WHT observing time proposal. All authors contributed to refining the text of the paper, or assisted in obtaining parts of the presented data set.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to K.W. (kw113@le.ac.uk)

METHODS

Linear and circular polarimetry. GRB 121024A triggered the Burst Alert Telescope (BAT) on board the Swift satellite at 02:56:12 UT on 24 October 2012 (Swift trigger 536580)¹⁹. We will use this trigger time as t_0 (that is, time since burst of an observation is $t = t_{\text{obs}} - t_0$) throughout. The prompt emission shows the burst probably belongs to the class of long bursts, with a duration $t_{90} = 69 \pm 32$ s (ref. 31). An X-ray and optical afterglow was found by the Swift X-ray telescope (XRT) and UV-optical telescope (UVOT)¹⁹. The redshift, $z = 2.298$, was found through afterglow spectroscopy with the X-shooter instrument on the Very Large Telescope³². Based on the initial brightness of the UVOT afterglow¹⁹ we activated our VLT polarimetry programme (programme 090.D-0789, PI Wiersema).

The data acquisition strategy, reduction and analysis closely follows that of our recent paper on GRB 091018³³, which in turn follows the specific recommendations set out for FORS polarimetry³³. We summarize those methods here, and give additional details specific to the case of GRB121024A.

Observations in the FORS2 R_{special} filter started at 9.2×10^3 s after burst. Following an acquisition image (Extended Data Fig. 1), polarimetry was acquired in imaging polarimetry (IPOL) mode. In this instrument mode, a super-achromatic half- or quarter-wavelength plate is used (for linear and circular polarimetry, respectively), after which a Wollaston prism is used to split the light into two beams, the so-called ordinary and extraordinary beam (hereafter the o and e beam) that have perpendicular polarization. These beams are imaged simultaneously: a slit mask is used to prevent overlap of the two beams on the chip (Extended Data Fig. 2). All linear polarization measurements are obtained using four rotation angles (0° , 22.5° , 45° and 67.5°) of the half-wavelength plate; for the circular polarimetry we used two angles (-45° and $+45^\circ$) of the quarter-wavelength plate. The GRB afterglow was positioned in the middle of the mask opening at the default (on-axis) position for FORS2 observations, on chip 1. A small number of linear polarization measurements were taken with a small dither (~ 16 pixels in the Y direction) to eliminate any potential effects of bad pixels/columns; none of these problems were apparent. Extended Data Tables 1 and 2 list the polarimetric observations and their resulting measurements.

Data reduction was done within IRAF, using bias and sky flat frames obtained on the same nights. Analysis of the data was done via aperture photometry, using IRAF scripts developed for this purpose, using the following method. We perform aperture photometry of all point sources present on both FORS2 chip 1 and 2, on the o and e images, using an aperture radius of 1.5 times the on-frame full-width at half-maximum (FWHM) of the point spread function (PSF). The PSF is found via a Gaussian fit on all point source objects, on a per-image basis, and is determined independently for the o and e beam as small differences in PSF shape may occur between the beams, particularly for objects far off-axis. The PSF value was used to set the aperture size; for example, the o beam on chip 1 is the weighted average of the FWHM of all point sources in that beam, chip and image. The sky subtraction was done using an annulus of inner and outer radii 3 and 4 times the FWHM, respectively. We only included sources for which the sky annulus was fully contained within the mask and showed no signs of saturation. Using this procedure we measured fluxes f_o and f_e for all point-like objects in all frames. Errors were determined using $\sigma_{\text{source}}^2 = g^{-1} \times f + (n_A + n_B/n_B) \sigma_B^2/\text{pixel}$, where g is the gain, f the flux (f_o and f_e , source minus the mean background) in the aperture, n_A and n_B the number of pixels in aperture and background region; and σ_B^2/pixel the variance per pixel in the sky level within the annulus.

The read noise contribution to the errors is negligible in these data. Images are not combined (that is, the depth of coverage per pixel is 1). The purpose of using the four angles for linear polarization, and two for the circular, is that the beams switch, eliminating several systematic errors (for example, flat field defects), resulting in increased reliability³³. We use the fluxes to derive the normalized flux difference at angle number i as $F_i = (f_{o,i} - f_{e,i})/(f_{o,i} + f_{e,i}) = (f_{o,i} - f_{e,i})/I$. We use the Stokes parameters (U , Q , V , I) to describe the source polarization state, often in normalized form (U/I , Q/I , V/I). These can be expressed in terms of F_i as $Q/I = \frac{2}{N} \sum_{i=0}^{N-1} F_i \cos(i\pi/2)$ and $U/I = \frac{2}{N} \sum_{i=0}^{N-1} F_i \sin(i\pi/2)$, where N is the number of half-wavelength plate positions (four positions in this case)^{17,33}. The circular-polarization Stokes parameter V/I is similarly computed as $V/I = \frac{1}{2} (F_{45} - F_{-45})$. The benefit of the multiple angle observations is clear from these equations. Further increasing the accuracy by observing at even more angles (for example, eight angles for linear polarization) is difficult for sources as faint as the afterglow of GRB 121024A—it would lead to observing times for single data points that are close to the polarization variability timescale.

We now proceed to correct for the linear polarization induced by dust in our own Galaxy which, along the line of sight to GRB 121024A, has $E(B - V) = 0.10$ (ref. 34). Using the relation $P_{\text{lin}} \leq 0.09 \times E(B - V)$ (ref. 35) this corresponds to a maximum induced polarization of $\sim 0.9\%$, which may be a noticeable fraction of the

detected polarization. To correct for Galactic-dust-induced polarization, we use the same methods as we used for GRB 091018 (ref. 17): we fit the field star Q , U distribution with a two-dimensional Gaussian function. We use three cuts on the sources entering the distribution: we require a polarimetric error $< 0.9\%$, a polarization value below 1.5% and a radial distance to the GRB position smaller than $2.5'$. This last cut is to counter the effects of instrumental polarization: the FORS2 instrument shows a broadly radial instrumental linear polarization pattern, with nil polarization on-axis (where the GRB is positioned), and values increasing with radial distance^{33,36,37}. By picking this cut-off we prevent the instrumental pattern from influencing the field objects' Q , U centroid determination too much, while still retaining enough sources for a reliable fit. We note that the number of bright field point sources within the mask is low in the case of GRB 121024A (87 data points from 9 individual stars enter the fit, considerably lower than in the case of GRB 091018; ref. 17), resulting in a somewhat larger uncertainty in the Galactic-dust-induced Q , U value. We find a value of $Q_{\text{Gal}} = -0.0020$, $U_{\text{Gal}} = -0.0046$. We use the standard deviation of the fitted 2D Gaussian as a measure for the uncertainty on the Galactic-dust-induced Stokes parameters, finding $\sigma_{Q_{\text{Gal}}} = 0.0040$, $\sigma_{U_{\text{Gal}}} = 0.0043$. We correct the afterglow Q , U values by subtracting the Galactic-dust-induced Q , U values. To facilitate comparison with models, we use the Stokes parameters to express the linear polarization in terms of the polarization degree P_{lin} and polarization angle θ as $P_{\text{lin}} = \sqrt{(Q^2 + U^2)}/I$ and $\theta = \frac{1}{2} \arctan(U/Q)$, where coordinates are chosen such that $\theta = 0^\circ$ for North, $\theta = 90^\circ$ for East. In the conversion from Q , U to P_{lin} , θ we account for the effects of polarization bias^{38,39} in the same manner as done for the data of GRB 091018^{17,33}. We correct the angles θ for the FORS2 instrumental zero angle offset³⁷. Similarly we can compute $P_{\text{cir}} = \sqrt{(V/I)^2}$: the sign of V/I gives the polarization direction (clockwise or anticlockwise). We note that both the instrument-induced circular polarization and linear to circular polarization crosstalk are well studied for the FORS instruments and far below our detection levels on the optical axis (where the afterglow is positioned)^{33,36,37}, and as such we expect no instrumental contribution to the detected V/I . A small amount of instrumental circular polarization is expected to be present for sources observed far off-axis^{36,37}, far from the GRB position. Since there are no bright (but unsaturated) stars near the GRB position, it is therefore not possible to use field sources as secure independent secondary standards^{33,36,37}. The circular polarization induced by scattering by the dust in our own Galaxy, with $E(B - V) = 0.10$ and the induced linear polarization values as above, does not contribute significantly to the observed P_{cir} (ref. 40).

In the second night of data, the afterglow of GRB 121024A shows a low level of linear polarization, which means that the uncertainty in the angle θ is relatively large, as $\sigma_\theta = \sigma_{P_{\text{lin}}}/2P_{\text{lin}}$. Nevertheless a clear difference in angle is visible with respect to the first night of data, consistent with a 90° change in polarization angle: the angle changes from $172^\circ \pm 2^\circ$ to $85^\circ \pm 10^\circ$ (Extended Data Table 1). The two first observations have angles somewhat discrepant from the pre-jet break average (Extended Data Table 1). Unfortunately we were not able to acquire infrared polarimetry simultaneous with the R -band polarimetry for scheduling reasons—simultaneous multi-wavelength polarimetry is the best way to directly measure polarization induced by scattering off dust particles in the host galaxy^{34,40,41}. However, the measured line-of-sight extinction and indeed the linear polarization light curve itself provide sufficient evidence that the host galaxy-dust-induced linear polarization must be small: the measured $A_V \approx 0.22$ mag gives a limit on the induced linear polarization of $< 0.7\%$ (assuming a Milky Way-like polarization curve³⁵).

Light curves and spectral energy distribution. The Swift XRT light curve is well described by three power-law segments, with parameters (errors at 90% confidence level) $\alpha_1 = 1.78^{+0.49}_{-0.10}$, $t_{\text{break},1} = 619^{+199}_{-348}$ s, $\alpha_2 = 0.83^{+0.13}_{-0.22}$, $t_{\text{break},2} = 3.4^{+1.5}_{-2.2} \times 10^4$ s and $\alpha_3 = 1.70^{+0.27}_{-0.36}$ (ref. 20). The uncertainty in the late break time is relatively large, as there are few data points post-break.

The optical afterglow of GRB 121024A was observed extensively with the seven channel (g' , r' , i' , z' , J , H , K_s filters) Gamma-Ray Burst Optical and Near-Infrared Detector (GROND⁴²) instrument, mounted on the ESO 2.2-m telescope at La Silla, Chile, providing photometric monitoring simultaneous to the polarimetry (Extended Data Table 3). These GROND observations will be described in more detail in a forthcoming, separate publication, but below we give some properties that are required to interpret the polarimetry. In addition to the GROND data, we observed the position of GRB 121024A with the ACAM instrument on the William Herschel Telescope (WHT) on 15 February 2013 (that is, 114.8 days after burst), with the aim of fixing the flux contribution of the host galaxy to the late-time light curve. We acquired 14×180 s exposures in Sloan r filter in relatively poor seeing conditions ($\sim 1.8''$). The resulting image (Extended Data Fig. 1) shows a clear detection of a source at the afterglow position, which we identify as the host galaxy of this GRB, confirming that a late-time detection in GROND data in g , r , i is dominated by the host. Its brightness, $r' = 24.03 \pm 0.20$, places the host at the bright end of the host galaxy luminosity distribution at this redshift⁴³. Taking this bright host galaxy magnitude into account, a light-curve break at around the same time as the break in XRT data is

apparent (Extended Data Fig. 3). We perform a joint fit to the XRT and GROND light curves (where only XRT data after the first XRT light-curve break, $t_{\text{break},1}$, are used) to constrain the break time. We fit using a model that consists of a smoothly broken power law, generally defined as

$$F_\nu(t) = F_\nu(t_{\text{break}}) \left(\left(\frac{t}{t_{\text{break}}} \right)^{\alpha_1 s} + \left(\frac{t}{t_{\text{break}}} \right)^{\alpha_2 s} \right)^{-1/s}$$

where t_{break} is the break time, α_1 and α_2 are pre- and post-break light-curve indices, and s is the break smoothness parameter. The pre- and post-break slopes and break smoothness are free parameters in our fit, and the break time is fixed to be the same for X-ray and optical/infrared wavelengths (that is, an achromatic break, and in addition a host galaxy contribution to the optical and infrared fluxes). This results in an acceptable fit statistic (reduced $\chi^2 = 157.38/132 = 1.192$), and the following parameters (uncertainties are 1σ): break time $t_{\text{break},2} = 3.72 \pm 0.07 \times 10^4$ s and break smoothness $s = 5.01 \pm 0.01$; pre-break light-curve decay indices $\alpha_{\text{pre,opt}} = 0.93 \pm 0.02$, $\alpha_{\text{pre,X-ray}} = 0.96 \pm 0.11$; post-break light-curve decay indices $\alpha_{\text{post,opt}} = 1.25 \pm 0.04$, $\alpha_{\text{post,X-ray}} = 1.67 \pm 0.10$. We identify this late, achromatic, break $t_{\text{break},2}$ with a so-called jet break. The resulting fit is shown in Extended Data Fig. 3. While in each optical band there are only a few data points post-break, GROND observes in seven bands simultaneously, making the break significant. The relatively shallow post-break optical decay is probably caused by the combination of bright host and smooth break: by the time the light curve asymptotes to its post-break index it is dominated by host galaxy light.

The line-of-sight extinction in the host galaxy and the spectral slopes are found by fitting an XRT+GROND spectral energy distribution (at time 11,085 s after trigger) with a SMC-like extinction law⁴⁴. The best fit, with reduced $\chi^2 = 1.04$, is obtained by a single power law (a broken power law does not result in significant fit improvement) with parameter $\beta = 0.88 \pm 0.01$, and a optical extinction in the V band of $A_V = 0.22 \pm 0.02$ mag.

The fact that X-ray and optical/infrared wavelengths have the same spectral index and that the pre-break decay indices are (within errors) identical, suggests that X-ray and optical are both in the same spectral regime, likely $\nu > \nu_c$. The achromatic nature of the light-curve break is consistent with a jet break interpretation, supporting our interpretation of the linear polarization behaviour of this afterglow.

Circular polarization of synchrotron emission. The linear and circular polarization degrees of the optically thin synchrotron emission from the electrons with a spectrum given by $N(\gamma_e, \alpha) = K\gamma_e^{-\beta}f(\alpha)$, where γ_e is the electron Lorentz factor, α is the electron pitch angle and K is the normalization factor, are given by^{11,45}

$$P_{\text{lin}} = \frac{p+1}{p+\frac{7}{3}}$$

$$P_{\text{cir}} = \frac{1}{\gamma_e} \frac{(2+p) \cot \theta + g(\theta)}{p} \frac{p+1}{p+\frac{7}{3}} \frac{\Gamma\left(\frac{3p+8}{12}\right) \Gamma\left(\frac{3p+4}{12}\right)}{\Gamma\left(\frac{3p+7}{12}\right) \Gamma\left(\frac{3p-1}{12}\right)}$$

Here θ is the viewing angle with respect to the magnetic field direction, Γ is the mathematical Gamma function, and we have defined

$$g(\theta) \equiv \frac{1}{f(\theta)} \left. \frac{df(\alpha)}{d\alpha} \right|_{\alpha=0}$$

These two equations are valid when $g(\theta) \ll \gamma_e$.

If the electron pitch-angle distribution is isotropic, that is, $g(\theta) = 0$, then $P_{\text{cir}} \sim \gamma_e^{-1}$. This simply means that the circular polarization contributions of electrons with pitch angles $\theta + \alpha$ and $\theta - \alpha$ nearly cancel out, and the remaining P_{cir} scales with the angular size of the beaming cone of the synchrotron emission, γ_e^{-1} .

The electrons with Lorentz factor γ_e mainly contribute to the synchrotron emission at frequency $\nu = \left(\frac{eB}{2\pi m_e c} \right) \gamma_e^2 \frac{\Gamma}{1+z}$, where the magnetic field strength B and the blast wave Lorentz factor Γ can be estimated by the standard external shock model²¹. Therefore, by calculating the Lorentz factor of the electrons producing the R-band emission, one can predict P_{cir} at the observing time as

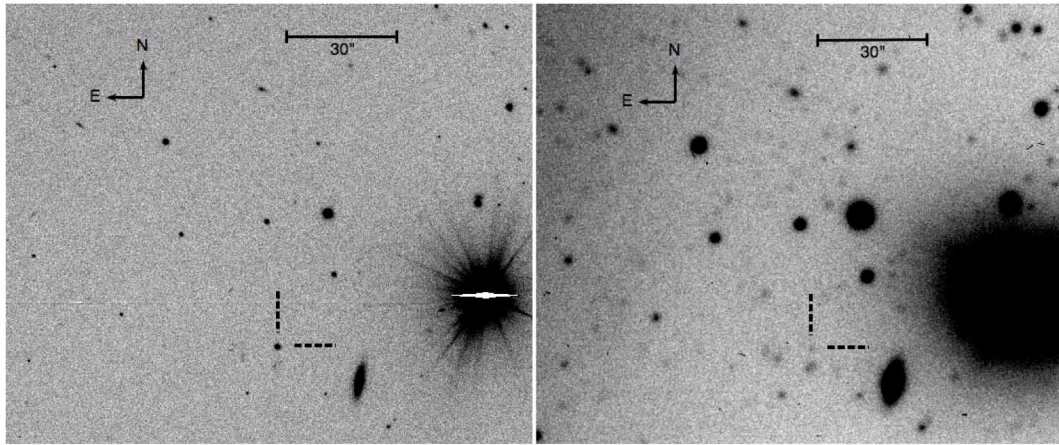
$$P_{\text{cir}} \approx 10^{-4} \epsilon_{B,-2}^{1/4} E_{52}^{1/8} n^{1/8}$$

where $\epsilon_B = 0.01 \epsilon_{B,-2}$ is the fraction of the magnetic energy density to the internal energy density, $E = 10^{52} E_{52}$ (erg) is the total blast wave energy and n (cm^{-3}) is the circumburst particle number density. This value does not strongly depend on the model parameters, and is very low (in spite of the assumption that the magnetic field is ordered) compared to the observed value of $P_{\text{cir}} = 0.61 \pm 0.13\%$. In reality, the magnetic field directions are largely random, as implied by the observed P_{lin} light curve. However, the linear and circular polarization degrees are reduced to the same extent by the randomness of the field, so the ratio $P_{\text{cir}}/P_{\text{lin}} \approx 10^{-4}$ is applicable also for the random field case, which is clearly inconsistent with the observed value $P_{\text{cir}}/P_{\text{lin}} \approx 0.15$.

In a situation where the pitch-angle distribution is not isotropic, the circular polarization contributions of electrons are not cancelled out and P_{cir} can be higher. The observed polarization ratio $P_{\text{cir}}/P_{\text{lin}}$, implies that $g(\theta)/\gamma_e \approx 0.1$ and then $g(\theta) \approx 10^3$, which means a highly anisotropic pitch-angle distribution.

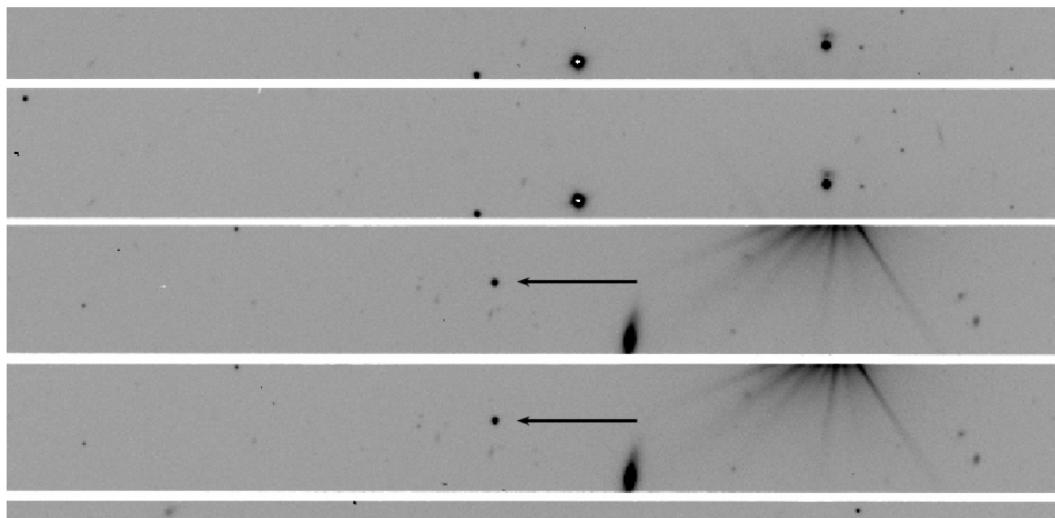
The detection of high circular polarization implies that the emitting plasma consists mainly of electrons and protons, rather than electrons and positrons, because the circular polarizations of the synchrotron emission of electrons and positrons perfectly cancel out⁴⁵. This implication is consistent with the emission model of the forward shock propagating in the circumburst medium.

31. Barthelmy, S. D. et al. GRB 121024A: Swift-BAT refined analysis. *GCN Circ.* **13889** (2012).
32. Tanvir, N. R. et al. GRB121024A: VLT/X-shooter redshift. *GCN Circ.* **13890** (2012).
33. Patat, F. & Romaniello, M. Error analysis for dual-beam optical linear polarimetry. *Publ. Astron. Soc. Pacif.* **118**, 146–161 (2006).
34. Schlegel, D. J., Finkbeiner, D. P. & Davis, M. Maps of dust infrared emission for use in estimation of reddening and cosmic microwave background radiation foregrounds. *Astrophys. J.* **500**, 525–553 (1998).
35. Serkowski, K., Matheson, D. S. & Ford, V. L. Wavelength dependence of interstellar polarisation and ratio of total to selective extinction. *Astrophys. J.* **196**, 261–290 (1975).
36. Bagnulo, S., Szeifert, T., Wade, G. A., Landstreet, J. D. & Mathys, G. Measuring magnetic fields of early-type stars with FORS1 at the VLT. *Astron. Astrophys.* **389**, 191–201 (2002).
37. FORS Users Manual Issue 91.1 (ESO, Doc. no. VLT-MAN-ESO-13100-1543, 2012); available at <http://www.eso.org/sci/facilities/paranal/instruments/fors/doc.html>.
38. Wardle, J. F. C. & Kronberg, P. P. The linear polarisation of quasi-stellar radio sources at 3.71 and 11.1 centimeters. *Astrophys. J.* **194**, 249–255 (1974).
39. Simmons, J. F. L. & Stewart, B. G. Point and interval estimation of the true unbiased degree of linear polarisation in the presence of low signal-to-noise ratios. *Astron. Astrophys.* **142**, 100–106 (1985).
40. Martin, P. G. Interstellar circular polarisation. *Mon. Not. R. Astron. Soc.* **159**, 179–190 (1972).
41. Klose, S. et al. Prospects for multiwavelength polarisation observations of GRB afterglows and the case GRB 030329. *Astron. Astrophys.* **420**, 899–903 (2004).
42. Greiner, J. et al. GROND — a 7-channel imager. *Publ. Astron. Soc. Pacif.* **120**, 405–424 (2008).
43. Hjorth, J. et al. The optically unbiased gamma-ray burst host (TOUGH) survey. I. Survey design and catalogs. *Astrophys. J.* **756**, 187–202 (2012).
44. Pei, Y. C. Interstellar dust from the Milky Way to the Magellanic Clouds. *Astrophys. J.* **395**, 130–139 (1992).
45. Melrose, D. B. *Non-thermal Processes in Diffuse Magnetised Plasmas* Vol. 1 (Gordon & Breach, 1980).



Extended Data Figure 1 | Host galaxy and afterglow image. Left, VLT FORS2 R_{special} -band acquisition image, with the afterglow indicated by broken

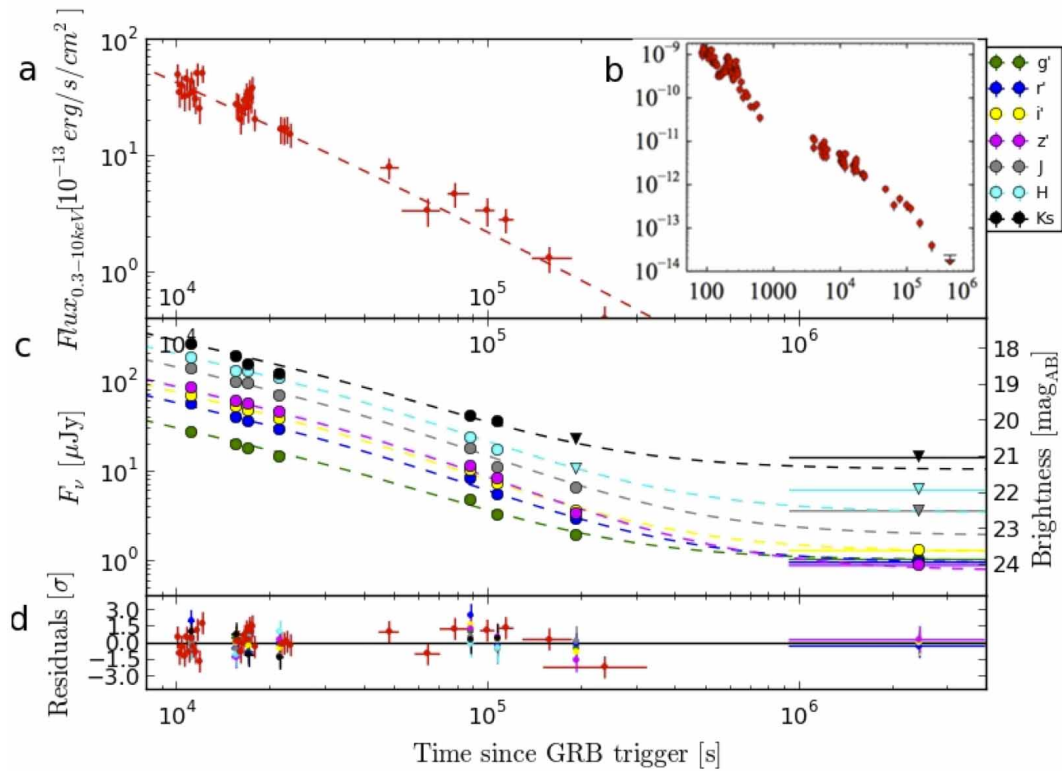
lines. Right, detection of the host galaxy in the late-time WHT ACAM r -band imaging.



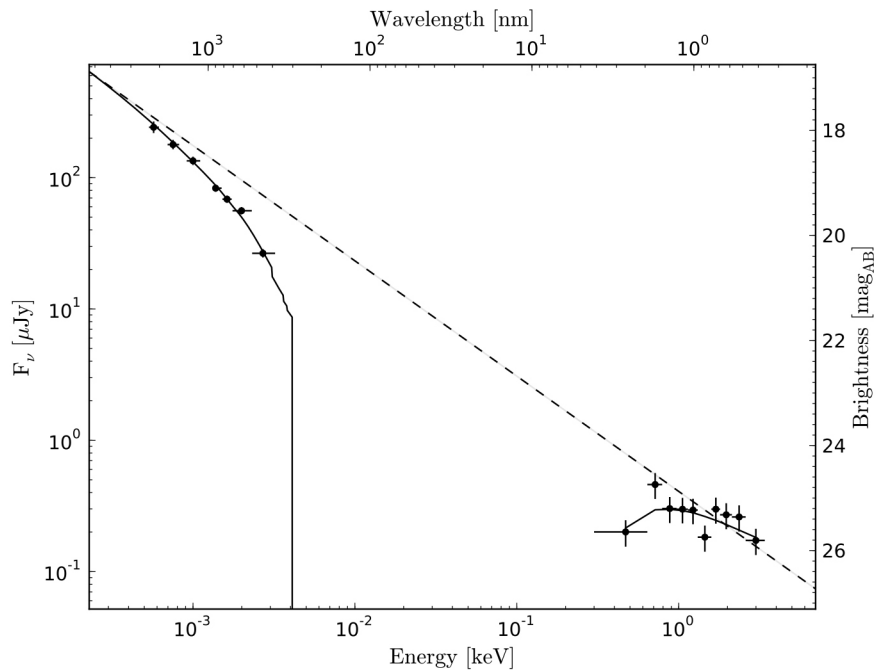
Extended Data Figure 2 | Polarimetry mask and afterglow brightness.

A small section of a single FORS2 R_{special} -band polarimetric exposure (this is the -45° angle chip 1 frame of the cir4 set), illustrating the shape of the aperture mask and brightness of the afterglow (indicated by an arrow). Each part of

the sky that falls into the open part of the rectangular mask is imaged twice, in perpendicular polarizations, the o and e beams, which is why the same objects each time appear in two non-overlapping strips.



Extended Data Figure 3 | Optical and X-ray afterglow light curves. **a**, Swift XRT X-ray light curve in the time span covered by GROND observations; **b**, the full XRT light curve. **c**, Full GROND light curves in all seven bands (key at top right). Overplotted in **a** and **c** is the best-fitting smoothly broken power law (Methods), with a host galaxy contribution to the optical data. Residuals of this fit are shown in **d**.



Extended Data Figure 4 | X-ray/optical spectral energy distribution of the afterglow of GRB 121024A. Shown is a spectral energy distribution using the seven GROND photometric bands and simultaneous Swift XRT X-ray data. The overplotted solid line is the best-fitting absorbed power law; the dashed line

shows the best-fitting power law without the effects of reddening and X-ray absorption. The horizontal error bars on the optical and X-ray data show the filter throughput and spectral bin-size, respectively. Vertical error bars show 1σ uncertainties on the fluxes.

Extended Data Table 1 | Linear polarimetry results

Datapoint identifier	Time since t_0 (mid; s)	P_{lin} (%)	θ (degrees)	Exposure time
lin1	9698	4.09 ± 0.20	163.7 ± 2.8	4×3 min
lin2	10638	4.83 ± 0.20	160.3 ± 2.3	4×3 min
lin3	14782	3.82 ± 0.20	182.7 ± 3.0	4×4 min
lin4	16042	3.12 ± 0.19	175.3 ± 3.5	4×5 min
lin5	17425	3.39 ± 0.18	178.0 ± 2.9	4×5 min
lin6	18841	3.49 ± 0.18	180.3 ± 3.0	4×5 min
lin7	20227	3.20 ± 0.18	174.5 ± 3.3	4×5 min
lin8	91624	0.34 ± 1.09	51.9 ± 67.5	$3 \times (4 \times 5)$ min
lin9	95828	1.49 ± 0.78	93.1 ± 26.6	$3 \times (4 \times 5)$ min
lin10	103028	2.66 ± 0.60	83.0 ± 12.6	$2 \times (4 \times 5)$ min
lin11	105814	0.86 ± 0.72	101.3 ± 36.4	$2 \times (4 \times 5)$ min

The polarization angle θ follows a standard coordinate convention: north = 0° , east = 90° . The values are corrected for Galactic-dust-induced polarization in Stokes parameter space, and polarization bias corrections are performed. Note that for the lin8–lin11 datapoints, we combine multiple exposure sets together: the polarization is low and the source faint.

Extended Data Table 2 | Circular polarimetry results

Datapoint identifier	Time since t_0 (mid; s)	V/I ($\times 100$)
cir1	11643	0.87 ± 0.21
cir2	12373	0.30 ± 0.21
cir3	13103	0.53 ± 0.22
cir4	13832	0.75 ± 0.23

Each datapoint consists of two exposures (-45° and $+45^\circ$ angles) of 5-min exposure time each. The sign of V/I is positive for all four data points (the sign distinguishes clockwise and anticlockwise circular polarization direction), and the circular polarization P_{cir} in per cent is therefore equal to the values in the third column. Uncertainties are 1σ .

Extended Data Table 3 | GROND optical and near-infrared photometry of the afterglow

Time since t_0 (mid; s)	g'	r'	i'	z'
11085.19	20.73 ± 0.08	19.80 ± 0.07	19.51 ± 0.07	19.25 ± 0.07
15496.95	21.09 ± 0.05	20.21 ± 0.07	19.84 ± 0.06	19.63 ± 0.07
17006.44	21.22 ± 0.05	20.31 ± 0.07	19.94 ± 0.06	19.68 ± 0.07
21430.12	21.45 ± 0.27	20.54 ± 0.07	20.20 ± 0.07	19.93 ± 0.09
106998.09	22.79 ± 0.24	21.99 ± 0.10	21.70 ± 0.12	21.48 ± 0.14
191934.13	23.31 ± 0.11	22.50 ± 0.09	22.13 ± 0.10	21.86 ± 0.13
2432398.02	24.24 ± 0.08	24.20 ± 0.09	23.80 ± 0.14	24.16 ± 0.15
Time since t_0 (mid; s)	J	H	K_s	
11085.19	18.67 ± 0.08	18.33 ± 0.09	17.96 ± 0.11	
15496.95	19.03 ± 0.07	18.70 ± 0.08	18.33 ± 0.10	
17006.44	19.07 ± 0.07	18.69 ± 0.08	18.56 ± 0.10	
21430.12	19.42 ± 0.08	18.90 ± 0.10	18.86 ± 0.13	
106998.09	20.99 ± 0.25	20.68 ± 0.30	20.29 ± 0.3	
191934.13	> 22.03	> 21.35	> 20.53	
2432398.02	> 22.50	> 21.92	> 21.02	

Magnitudes are as observed; uncertainties are 1σ ; upper limits are 3σ .

Classical shear cracks drive the onset of dry frictional motion

Ilya Svetlizky¹ & Jay Fineberg¹

Frictional processes entail the rupture^{1,2} of the ensemble of discrete contacts defining a frictional interface^{3,4}. There are a variety of views on how best to describe the onset of dry frictional motion. These range from modelling friction with a single degree of freedom, a ‘friction coefficient’^{3,5}, to theoretical treatments using dynamic fracture^{5–8} to account for spatial and temporal dynamics along the interface. We investigated the onset of dry frictional motion by performing simultaneous high-speed measurements of the real contact area and the strain fields in the region surrounding propagating rupture tips within the dry (nominally flat) rough interfaces formed by brittle polymer blocks. Here we show that the transition from ‘static’ to ‘dynamic’ friction is quantitatively described by classical singular solutions for the motion of a rapid shear crack^{5,9–13}. We find that these singular solutions, originally derived to describe brittle fracture, are in excellent agreement with the experiments for slow propagation, whereas some significant discrepancies arise as the rupture velocity approaches the Rayleigh wave speed. In addition, the energy dissipated in the fracture of the contacts remains nearly constant throughout the entire range in which the rupture velocity is less than the Rayleigh wave speed, whereas the size of the dissipative zone undergoes a Lorentz-like contraction as the rupture velocity approaches the Rayleigh wave speed. This coupling between friction and fracture is critical to our fundamental understanding of frictional motion and related processes, such as earthquake dynamics.

A dry frictional interface is composed of an ensemble of discrete contacts whose total area (the real contact area, A) is orders of magnitude smaller than the nominal contact area⁴. Local motion (slip) is initiated when contacts are broken via propagating ruptures¹. Characterization of the dynamic fields that drive these ruptures and how they couple to the dissipative mechanisms on the interface are therefore critical to our fundamental understanding of frictional motion. These fields also describe the elastic radiation emitted by the rupture whose form is important to interpretation of near-field seismic signals radiated by earthquakes^{14,15}.

Linear elastic fracture mechanics¹⁰ (LEFM) is the fundamental theoretical framework describing brittle fracture. LEFM predicts a universal singular form of stresses (and strains) in the vicinity of a crack tip. These singular stresses are only regularized by dissipative^{10–13} processes that take place as this singular limit is approached. Shear-driven (mode II) fracture is generally considered to be impossible in bulk isotropic materials⁹ because it is believed that a crack will rotate under imposed shear so as to fracture under pure tension. An exception to this may occur for highly anisotropic conditions, as in the case of the weak interface plane that defines a frictional interface^{11,16–18}. Examples of such systems range from natural faults to single adhesive contacts. Although both contact separation⁸ and earthquake dynamics^{5,7,12,13,15} have long been theoretically described using fracture mechanics, there has been no direct experimental evidence that quantitative predictions of LEFM, such as the universal singular forms of stresses and material motion at tips of rapidly moving ruptures, really occur in frictional failure. We will demonstrate that, in the brittle material studied, frictional ruptures along dry

and (nominally flat) rough interfaces can indeed be quantitatively described by LEFM.

We study the structure of the elastic fields driving ruptures that propagate within an interface formed between two initially flat poly(methyl-methacrylate) blocks, which were uniformly roughened to a 3- μm r.m.s. roughness. The blocks are pressed together with an externally imposed normal force, $F_N = 5,500\text{ N}$ (5 MPa of nominal pressure). Once F_N is applied, a shear force, F_S , is quasi-statically incremented until stick-slip motion initiates (Fig. 1a). The complete two-dimensional strain tensor, $\varepsilon_{ij}(t)$, is measured at 19 locations along and $\sim 3.5\text{ mm}$ above the interface at 1,000,000 samples per second (Fig. 1a). We simultaneously measure the displacement parallel to the interface, u_x , at specific locations adjacent to the strain gages. At the same time, we measure the real contact area, $A(x,t)$, with high spatial resolution at 580,000 frames per second (see Methods). We focus on the short time intervals bracketing slip initiation.

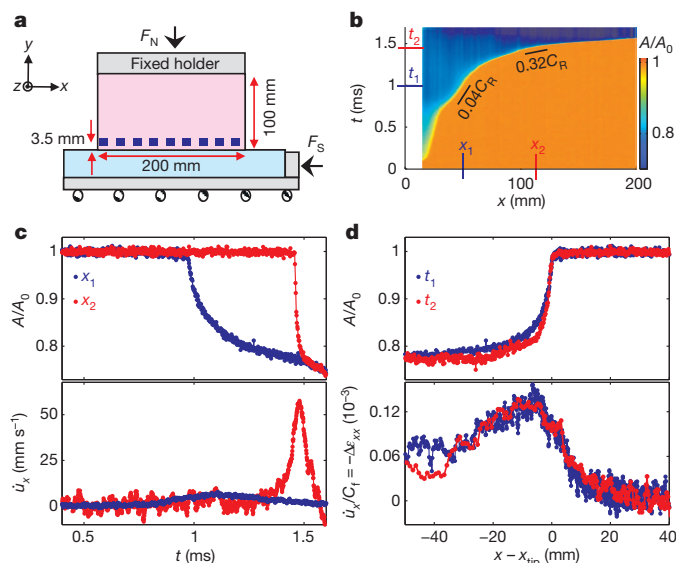


Figure 1 | Experimental system and slow rupture scaling. **a**, 19 Rosette strain gauges (blue squares) are mounted 3.5 mm above the frictional interface. The three strain components were measured simultaneously every 1 μs . **b**, The evolution of real contact area, $A(x,t)$ (normalized at nucleation time, $t = 0$), along the 200-mm quasi-one-dimensional interface of a slow rupture front, which nucleated at $x = 0$ and accelerated to $0.3C_R$. **c**, $A(t)$ (top) and $\dot{u}_x(t)$ (bottom) (\dot{u}_x is the x component of the displacement field measured 3.5 mm above the interface) at locations x_1 (blue dots) and x_2 (red dots) denoted in **b**. **d**, Collapse of both $A(x,t)$ (top) and \dot{u}_x/C_f (bottom) is obtained by plotting the data as a function of the distance $x - x_{tip}$ when the rupture tip, x_{tip} , arrives at x_1 (blue, at time t_1) and x_2 (red, at time t_2). C_f is the local front velocity. $A(x, t_i)$ was measured directly whereas the spatially dependent $\dot{u}_x(x)$ was constructed from the time series $\dot{u}_x(t)$ using $\dot{u}_x(x, t) = \dot{u}_x(x - C_f t)$. Variations of $C_f(x)$ are accounted for as described in Methods. During the front's passage, $-\dot{u}_x/C_f$ corresponds to changes of the strain component parallel to the interface, $\Delta\varepsilon_{xx}$.

¹The Racah Institute of Physics, The Hebrew University of Jerusalem, Givat Ram, Jerusalem 91904, Israel.

We define the location of a rupture front, x_{tip} , as the point where the initial value of $A(x, t)$ is reduced by 3%. x_{tip} both represents the fracture of contacts and signals the initiation of local slip². For a steadily moving front, $\varepsilon_{ij}(x, t) = \varepsilon_{ij}(x - C_f t)$, where C_f is the rupture front velocity. When C_f varies slowly, $\varepsilon_{ij}(x, t) = \varepsilon_{ij}(x - \int C_f dt)$, enabling conversion from temporal $\varepsilon_{ij}(t)$ to spatial measurements $\varepsilon_{ij}(x - x_{\text{tip}})$, using $C_f(t)$ obtained from $A(x, t)$ measurements (see Methods). For each front passage we define strain tensor variations, $\Delta\varepsilon_{ij}$, by subtracting the initial strains from ε_{xx} and ε_{yy} and residual strain from ε_{xy} .

We present a typical measurement of $A(x, t)$ for a slow, but accelerating, rupture front in Fig. 1b. C_f is well below the Rayleigh wave speed, C_R ($\sim 1,255 \text{ m s}^{-1}$) over the entire duration of this experiment. As C_f increases, the timescale of the weakening process (drop of $A(x, t)$) shortens (Fig. 1c, top) and material element velocities, $\dot{u}_x(u)$ is the displacement field), increase significantly (Fig. 1c, bottom). In contrast to $A(x, t)$, which is an interfacial property, \dot{u}_x constitutes a lower bound of the slip velocity at the interface, because it is measured 3.5 mm from the interface (see Methods). When plotted in the reference frame of the moving rupture tip, $x - x_{\text{tip}}$, the $A(x, t)$ collapse (Fig. 1d, top) to a single curve. Furthermore, $\dot{u}_x(x)$, when scaled by C_f , collapses to a well-defined functional form (Fig. 1d, bottom) corresponding to $\Delta\varepsilon_{xx} = -\dot{u}_x(x)/C_f$ (Extended Data Fig. 2). Note that contacts are broken for $x - x_{\text{tip}} < 0$ and unbroken when $x - x_{\text{tip}} > 0$.

Figure 2a demonstrates that for slow ruptures ($\sim 0.01C_R < C_f < \sim 0.3C_R$) the measured variations of each strain tensor component, $\Delta\varepsilon_{ij}$, collapse to well-defined functional forms. This breaks down as C_f increases; all $\Delta\varepsilon_{ij}$ amplitudes grow significantly while strain oscillations

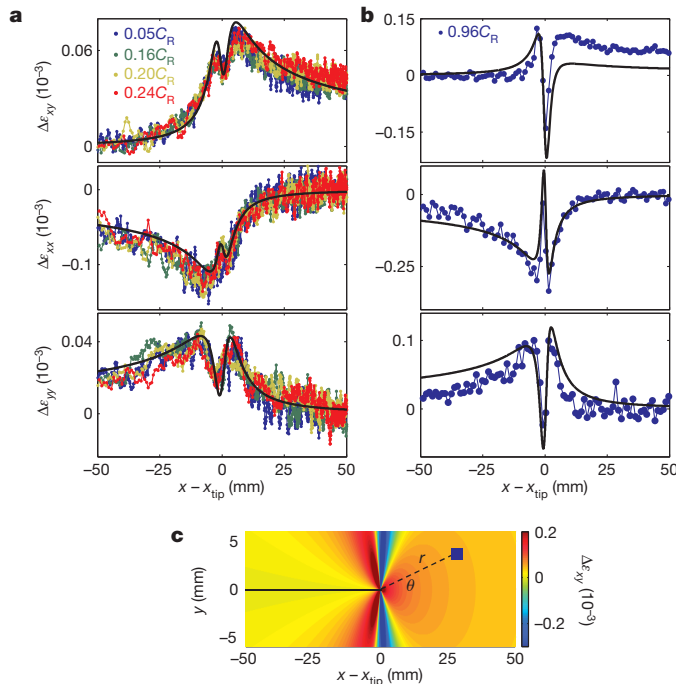


Figure 2 | The functional form of the elastic strains. **a**, Measurements of strain tensor variations, $\Delta\varepsilon_{ij}$, after subtracting the initial strain from ε_{xx} and ε_{yy} and residual strain from ε_{xy} . Colours represent different slow front velocities ($0.05C_R < C_f < 0.24C_R$; see key). Corresponding LEFM predictions (equation (1)) for plane strain boundary conditions are plotted in black ($C_f = 0.2C_R$). The fracture energy $\Gamma \approx 1.1 \text{ J m}^{-2}$ is the sole free parameter. **b**, The mean value of $\Delta\varepsilon_{ij}$ for three typical measurements at $C_f \approx 0.96C_R$. Equation (1) is plotted (black line) for the same Γ as in **a**, and still describes the larger amplitudes and strong oscillations of rapidly propagating ruptures. **c**, Shear strain variations, $\Delta\varepsilon_{xy}$, surrounding the rupture tip predicted by equation (1) for $C_f = 0.96C_R$. The blue square denotes the strain gauge location relative to the approaching rupture tip. Note the strong angular dependence that drives the violent oscillations in **b**.

are strongly amplified. Typical measurements of $\Delta\varepsilon_{ij}$ for high velocities ($0.96C_R$) are presented in Fig. 2b.

We now compare the measured strains, ε_{ij} , to the universal asymptotic solutions for mode II cracks predicted by LEFM. In our experiment, external tensile stresses ($\sigma_{xx}^0, \sigma_{yy}^0$) exist and the crack faces are subjected to a residual frictional shear stress, τ^r , so that the stresses can be written as $\sigma_{ij} = \Delta\sigma_{ij} + \begin{pmatrix} \sigma_{xx}^0 & \tau^r \\ \tau^r & \sigma_{yy}^0 \end{pmatrix}$. Owing to the linearity of the governing equations, the stress field variations, $\Delta\sigma_{ij}$, can be mapped¹¹ to the stress-free conditions that define the mode II crack problem. Therefore, near the rupture tip, $\Delta\sigma_{ij}$ have the singular form:

$$\Delta\sigma_{ij} = \frac{K_{II}}{\sqrt{2\pi r}} \Sigma_{ij}^{II}(\theta, C_f) \quad (1)$$

where (r, θ) are polar coordinates relative to the crack tip, $\Sigma_{ij}^{II}(\theta, C_f)$ is a known universal function and K_{II} is a scalar coefficient known as the stress intensity factor¹⁰. $\Delta\sigma_{ij}$ are linearly related to the measured strain variations $\Delta\varepsilon_{ij}$ via the elastic moduli and the use of plane strain boundary conditions (see Methods). Because the strain gauges are displaced from the interface, each $\Delta\varepsilon_{ij}(x)$ measurement involves both radial ($50 \text{ mm} > r > 3.5 \text{ mm}$) and angular ($0 < \theta < \pi$) variations (see Fig. 2c).

In Fig. 2a we show that for slow ruptures all of the measured strain components $\Delta\varepsilon_{ij}$ agree well with the strains corresponding to equation (1) (black lines), where the only free parameter is K_{II} . The apparent data collapse in Fig. 2a occurs because these LEFM solutions for $C_f < 0.4C_R$ are indistinguishable. K_{II} is related to G , the energy flux per unit area (the energy release rate), via¹⁰:

$$G = \frac{(1 - \nu^2)}{E} f_{II}(C_f) K_{II}^2 \quad (2)$$

where $f_{II}(C_f)$ is a known universal function that is fairly constant for low velocities and diverges as $C_f \rightarrow C_R$. When G is balanced by the fracture energy Γ (the energy dissipated per unit area), equation (2) uniquely relates K_{II} for each C_f to Γ . Using equation (1) to measure K_{II} , the energy balance condition ($G = \Gamma$) therefore yields a single value $\Gamma \approx 1.1 \pm 0.3 \text{ J m}^{-2}$ for the LEFM predictions presented in Fig. 2a over $0.01C_R < C_f < \sim 0.3C_R$.

A typical example of rupture at high C_f (Fig. 2b) demonstrates that for the same constant value of Γ most characteristic features of $\Delta\varepsilon_{ij}$ observed at higher rupture velocities are also well described by equation (1). For example, the violent high-amplitude strain oscillations that occur when the rupture tip passes beneath the measurement point (Fig. 2c) are due to the singular nature of $\Sigma_{ij}^{II}(\theta, C_f)$. Figure 2b, top, however, demonstrates that equation (1) clearly fails to describe $\Delta\varepsilon_{xy}$ for $x - x_{\text{tip}} > 0$.

To systematically study how $\Delta\varepsilon_{ij}$ varies over $0.01 < C_f < 0.99C_R$, we characterize $\Delta\varepsilon_{ij}$ variations using $\delta\varepsilon_{xx}$, $\delta\varepsilon_{yy}$, $\delta\varepsilon_{xy}$ and $\delta\varepsilon_{xy}^{\text{Osc}}$, as defined in the insets within Fig. 3. With the only input being the constant value of Γ measured at low velocities, we see that, for all C_f , the theory quantitatively describes both the strain variations $\delta\varepsilon_{xx}$ and $\delta\varepsilon_{yy}$ (Fig. 3a) and the violent high-amplitude strain oscillations, $\delta\varepsilon_{xy}^{\text{Osc}}$, (Fig. 3b, top) evident in the example in Fig. 2b.

Although $\delta\varepsilon_{xy}$ is also well-described by the theory for $C_f/C_R < 0.4$, we find (Fig. 3b, bottom) that a systematic failure of the classic universal form predicted by LEFM (equation (1)) occurs, as seen in Fig. 2b, top. These deviations are first observed at $C_f \approx 0.7C_R$ and increase dramatically as $C_f \rightarrow C_R$; experiments reveal a sharp and systematic increase of $\delta\varepsilon_{xy}$, whereas the theory predicts vanishingly small values as $C_f \rightarrow C_R$. Note that for $C_f/C_R < 0.4$ the apparent collapse of all strain components (Fig. 2a) is a natural consequence of the theory.

We now consider the value of Γ . Whereas the bulk fracture energy, $\Gamma_{\text{bulk}} = 2,000 \text{ J m}^{-2}$, is a material property of poly(methylmethacrylate), the interface value, Γ , is directly related to $A(x, t)$, which is determined by F_N . It is interesting that the measured value $\Gamma \approx 1.1 \text{ J m}^{-2}$ for our experimental conditions is approximately Γ_{bulk} when the sparseness of the contacts⁴ ($A \ll A_{\text{nominal}}$) is accounted for (see Methods).

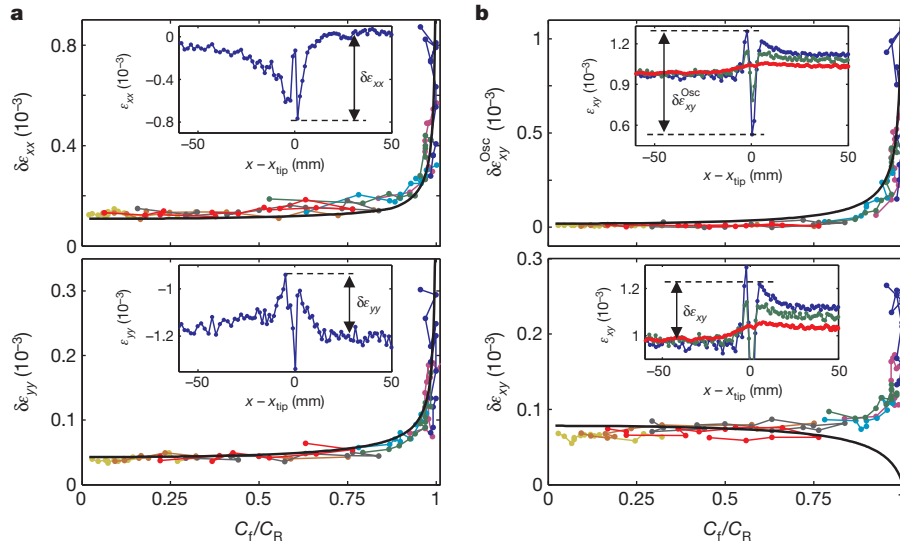


Figure 3 | Systematic comparison of measured strains to LEFM.

a, Measured elastic strains, ε_{xx} (top, inset) and ε_{yy} (bottom, inset), are characterized by their signal amplitudes, $\delta\varepsilon_{xx}$ and $\delta\varepsilon_{yy}$ (black arrows). Main panels: $\delta\varepsilon_{xx}$ (top) and $\delta\varepsilon_{yy}$ (bottom) are plotted for $0.01C_R < C_f < 0.99C_R$. **b**, We characterize ε_{xy} by the oscillatory amplitude, $\delta\varepsilon_{xy}^{\text{Osc}}$ (top, inset) and the shear strain drop, $\delta\varepsilon_{xy}$ (bottom, inset). **a**, **b**, $C_R = 1,255 \pm 10 \text{ m s}^{-1}$, plotted

strains are in units of 10^{-3} . Colours represent different stick-slip sequences, where different measurement locations are connected by a solid line. Classic LEFM predictions, for constant $\Gamma \approx 1.1 \text{ J m}^{-2}$, are plotted in black. Equation (1) successfully describes $\delta\varepsilon_{xx}$, $\delta\varepsilon_{yy}$ (**a**) and $\delta\varepsilon_{xy}^{\text{Osc}}$ (**b**, top) while systematically failing to describe $\delta\varepsilon_{xy}$ (**b**, bottom).

This suggests that significant plastic deformation (the major contribution to Γ_{bulk}) may also take place within the contacting asperities, as surmised in ref. 2.

Although LEFM enables us to measure Γ , it does not address the details of the dissipative region. As $A(x, t)$ is measured on the interface, it characterizes cohesive zone (interface weakening region) properties. Figure 4a reveals that the contact area reduction, ΔA , increases with C_f , suggesting that corrections to our assumption of constant Γ may be required at high velocities. Γ could indeed vary significantly in the singular ($C_f \approx C_R$) region of $\Delta\varepsilon_{ij}$ while still being consistent with the data, as when approaching the singularity our finite ($\pm 10 \text{ m s}^{-1}$) measurement accuracy of C_f and C_R limits our ability to resolve Γ .

Although the non-monotonic behaviour of A with $x - x_{\text{tip}}$ suggests interesting dynamics as $C_f \rightarrow C_R$, for reasons of simplicity we characterize the scale over which $A(x, t)$ is reduced by single length scale, X_c (Fig. 4a, bottom). We suggest that X_c provides a direct measure of the cohesive zone size. Figure 4b demonstrates that X_c is not constant, but systematically contracts with increasing C_f . As $C_f \rightarrow C_R$, X_c appears to

approach zero. Fracture mechanics predicts such an effective ‘Lorentz contraction’ of all length scales in the propagation direction^{5,10,12}. In particular, we expect $X_c = X_c^0/f_{\text{II}}(C_f)$, where $f_{\text{II}}(C_f)$ is the universal function (equation (2)) relating K_{II} to G , which diverges at C_R and $X_c^0 = X_c(C_f = 0)$. Figure 4b (black line) demonstrates that this general result of elastodynamic theory describes the measurements well.

Figure 4b indicates that in the slow rupture regime, strain measurements were performed at scales of $\sim X_c$, which characterizes the scale of the cohesive zone. The universal form, predicted by equation (1), is expected to describe the elastic fields at distances far from the cohesive zone, but still small compared to other dimensions of the system¹¹. The good agreement with LEFM (Fig. 2a) is, nevertheless, consistent (see Extended Data Fig. 4) with simple cohesive zone models coupled to LEFM¹¹.

In this light, as X_c contracts by nearly an order of magnitude as $C_f \rightarrow C_R$, we should expect equation (1) to become increasingly more accurate in precisely the velocity regime where it performs badly (compare $\delta\varepsilon_{xy}$ in Fig. 3b, bottom). $\delta\varepsilon_{xy}(C_f)$, as Fig. 3b, inset, demonstrates, correlates with the initially imposed shear stresses ahead of the front. These initial stresses have been shown to govern rupture velocity selection^{19–21}. Higher order contributions to this LEFM solution (see Extended Data Fig. 5) do not correct these discrepancies. As ruptures where $C_f \rightarrow C_R$ are commonplace in both frictional failure and earthquake dynamics, understanding this failure of the classical LEFM solution is important; the form of the singular stress fields surrounding rapid ruptures has significant ramifications for earthquake branching and fault bifurcation²².

Knowledge of Γ and X_c enables us to estimate elusive but long-sought^{13,16–18} dynamical quantities at the interface that include peak stresses and slip velocities. The contraction of X_c tells us (see Methods) that quantities measured off-interface increasingly deviate from their interface values as $C_f \rightarrow C_R$; for example, the peak value (strain = 0.8×10^{-3}) of $\delta\varepsilon_{xx}$ ($= -\dot{u}_x/C_f$), which corresponds to $\dot{u}_x \approx 1 \text{ m s}^{-1}$, can be an order of magnitude higher on the interface¹⁸.

Whereas we have demonstrated that equations (1) and (2) provide an excellent description of the fields that drive frictional motion in a specific material under dry conditions, we expect the results to be generally valid as long as the assumptions underpinning the LEFM solution are obeyed. These are: (1) the dissipative scale is small compared to the size of the elastic region in which the stress singularity

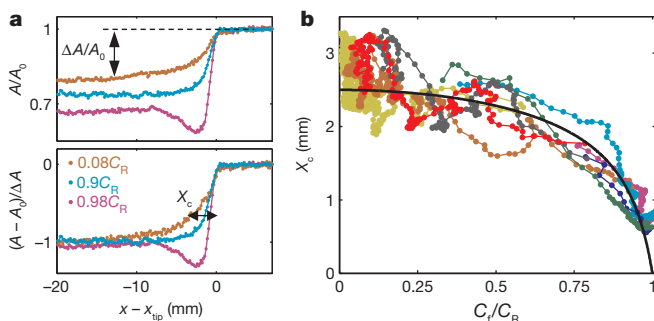


Figure 4 | ‘Lorentz’ contraction of the cohesive zone. **a**, Top; the contact area reduction, $\Delta A/A_0$, increases for faster ruptures and non-monotonic behaviour is observed for $C_f > 0.9C_R$. Bottom, we characterize the weakening region (the reduction of A) by X_c , the length scale where a 60% reduction of $\Delta A/A_0$ occurs. Contraction of X_c with C_f is apparent. **b**, X_c as a function of C_f . Colours represent the different slip events in Fig. 3. Black line: LEFM prediction, $X_c = X_c^0/f_{\text{II}}(C_f)$, where $X_c^0 = X_c(C_f \approx 0)$. $X_c(C_f \approx C_R) < 0.5 \text{ mm}$ are limited by our temporal resolution.

develops; and (2) the residual stress, τ^r , remains fairly constant. So long as the dissipative region is sufficiently confined, its precise nature (for example, whether granular, powdered or consisting of contacting asperities) only contributes to the value of $\Gamma(C_f)$. This description may have to be adapted for lubricated boundaries where τ^r may be strongly slip-rate dependent⁶, or even break down within highly damaged regions whose spatial extent is large compared to the surrounding elastic material.

The measurements in Figs 2 and 3 may therefore provide a precise quantitative framework for describing frictional motion along rough, nominally flat, surfaces. Such high-speed strain measurements¹⁴ are rarely performed adjacent to natural faults. It is interesting to consider if the universal description of strains observed here is also valid during earthquakes. If so, high-speed single-point measurements of $\Delta\epsilon_{ij}$ adjacent to faults could provide sensitive measurements of both earthquake velocities and Γ . These measurements could reveal rupture-velocity dependence of both intrinsic values of Γ and values resulting from near-fault damage. These measurements could also clearly differentiate between different types of fracture modes such as supershear¹⁵ or slow ruptures, which are increasingly observed in both experiments^{1,23–25} and natural faults²⁶. Our results indicate that the structure of slow ruptures is characterized by LEFM. The mechanisms governing their appearance are, however, not yet clear and may depend on properties of the underlying friction law (for example, a velocity strengthening range)^{27,28}.

Let us now consider the threshold for frictional failure. Recent studies^{14,29,30} have shown that the static friction coefficient is not a material property, but depends on the applied loading. The failure of bulk materials is described by the Griffith condition—the critical energy release rate, G_{crit} , being equal to Γ . We have shown that interface rupture propagation is described by, essentially, the same framework. This suggests that an analogous ‘Griffith’ condition may exist for the onset of rupture propagation for a frictional interface, and by incorporating applied loading conditions it may be possible to calculate G_{crit} .

METHODS SUMMARY

The Methods section contains additional material about sample preparation and characterization together with detailed descriptions of the loading system, real contact area measurements, slip and strain measurements. Also presented is a comparison between direct measurements of the material velocities and derived values obtained using the relation $\delta\epsilon_{xx} = -\dot{u}_x/C_f$. More detailed descriptions of the analysis techniques (comparison of LEFM to measured strains, converting temporal to spatial measurements) used in the manuscript are also presented. Included is additional information that demonstrates that both higher-order (non-singular) strain field terms and the existence of a (simple) cohesive zone of size X_c have negligible effects on the comparisons of equation (1) to the data measured in our experiments. We additionally include a brief section that shows how estimates of cohesive zone properties (peak stress, interface slip velocity, and weakening distance on the interface) can be obtained from our measurements of Γ and X_c . We also describe how we relate the measured value of Γ to measured bulk values, Γ_{bulk} , of the fracture energy.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 18 November 2013; accepted 25 February 2014.

1. Rubinstein, S. M., Cohen, G. & Fineberg, J. Detachment fronts and the onset of dynamic friction. *Nature* **430**, 1005–1009 (2004).
2. Ben-David, O., Rubinstein, S. M. & Fineberg, J. Slip-stick and the evolution of frictional strength. *Nature* **463**, 76–79 (2010).

3. Bowden, F. P. & Tabor, D. *The Friction and Lubrication of Solids* 2nd edn (Oxford Univ. Press, 2001).
4. Dieterich, J. H. & Kilgore, B. D. Imaging surface contacts: power law contact distributions and contact stresses in quartz, calcite, glass and acrylic plastic. *Tectonophysics* **256**, 219–239 (1996).
5. Scholz, C. H. *The Mechanics of Earthquakes and Faulting* 2nd edn (Cambridge Univ. Press, 2002).
6. Brener, E. A. & Marchenko, V. I. Frictional shear cracks. *JETP Lett.* **76**, 211–214 (2002).
7. Ben-Zion, Y. Collective behavior of earthquakes and faults: continuum-discrete transitions, progressive evolutionary changes, and different dynamic regimes. *Rev. Geophys.* **46**, RG4006 (2008).
8. Johnson, K. L. Adhesion and friction between a smooth elastic spherical asperity and a plane surface. *Proc. R. Soc. Lond. A* **453**, 163–179 (1997).
9. Fineberg, J. & Marder, M. Instability in dynamic fracture. *Phys. Rep.* **313**, 1–108 (1999).
10. Freund, L. B. *Dynamic Fracture Mechanics* (Cambridge, 1990).
11. Palmer, A. C. & Rice, J. R. The growth of slip surfaces in the progressive failure of over-consolidated clay. *Proc. R. Soc. Lond. A* **332**, 527–548 (1973).
12. Rice, J. R. in *Physics of the Earth's Interior* (eds Dziewonski, A. M. & Boschi, E.) 555–649 (Italian Physical Society and North-Holland, 1980).
13. Ida, Y. Cohesive force across tip of a longitudinal-shear crack and Griffiths specific surface energy. *J. Geophys. Res.* **77**, 3796–3805 (1972).
14. Borchardt, R. D., Johnston, M. J. S., Glassmoyer, G. & Dietel, C. Recordings of the 2004 Parkfield earthquake on the General Earthquake Observation System array: implications for earthquake precursors, fault rupture, and coseismic strain changes. *Bull. Seismol. Soc. Am.* **96**, S73–S89 (2006).
15. Dunham, E. M. & Archuleta, R. J. Evidence for a supershear transient during the 2002 Denali fault earthquake. *Bull. Seismol. Soc. Am.* **94**, S256–S268 (2004).
16. Okubo, P. G. & Dieterich, J. H. Effects of physical fault properties on frictional instabilities produced on simulated faults. *J. Geophys. Res.* **89**, 5817–5827 (1984).
17. Ohnaka, M. *The Physics of Rock Failure and Earthquakes* (Cambridge Univ. Press, 2013).
18. Lykotrafitis, G., Rosakis, A. J. & Ravichandran, G. Self-healing pulse-like shear ruptures in the laboratory. *Science* **313**, 1765–1768 (2006).
19. Ben-David, O., Cohen, G. & Fineberg, J. The dynamics of the onset of frictional slip. *Science* **330**, 211–214 (2010).
20. Trømborg, J., Scheibert, J., Amundsen, D. S., Thøgersen, K. & Mølle-Sørensen, A. Transition from static to kinetic friction: insights from a 2D model. *Phys. Rev. Lett.* **107**, 074301 (2011).
21. Kammer, D. S., Yastrebov, V. A., Spijker, P. & Molinari, J. F. On the propagation of slip fronts at frictional interfaces. *Tribol. Lett.* **48**, 27–32 (2012).
22. Poliakov, A. N. B., Dmowska, R. & Rice, J. R. Dynamic shear rupture interactions with fault bends and off-axis secondary faulting. *J. Geophys. Res.* **107**(B11), 2295 (2002).
23. Nielsen, S., Taddeucci, J. & Vinciguerra, S. Experimental observation of stick-slip instability fronts. *Geophys. J. Int.* **180**, 697–702 (2010).
24. Latour, S. *et al.* Ultrafast ultrasonic imaging of dynamic sliding friction in soft solids: the slow slip and the super-shear regimes. *Europhys. Lett.* **96**, 59003 (2011).
25. Yamaguchi, T. *et al.* Gutenberg-Richter's law in sliding friction of gels. *J. Geophys. Res.* **116**, B12306 (2011).
26. Beroza, G. C. & Ide, S. Slow earthquakes and nonvolcanic tremor. *Annu. Rev. Earth Planet. Sci.* **39**, 271–296 (2011).
27. Bar Sinai, Y., Brener, E. A. & Bouchbinder, E. Slow rupture of frictional interfaces. *Geophys. Res. Lett.* **39**, L03308 (2012).
28. Kaproth, B. M. & Marone, C. Slow earthquakes, preseismic velocity changes, and the origin of slow frictional stick-slip. *Science* **341**, 1229–1232 (2013).
29. Passelègue, F. X., Schubnel, A., Nielsen, S., Bhat, H. S. & Madariaga, R. From sub-Rayleigh to supershear ruptures during stick-slip experiments on crustal rocks. *Science* **340**, 1208–1211 (2013).
30. Ben-David, O. & Fineberg, J. Static friction coefficient is not a material constant. *Phys. Rev. Lett.* **106**, 254301 (2011).

Acknowledgements This work was supported by the James S. McDonnell Fund, the European Research Council (grant no. 267256) and the Israel Science Foundation (grant 76/11). We thank E. Bouchbinder and G. Cohen for comments. We especially thank M. Adda-Bedia for insights that helped us understand the data's relation to LEFM.

Author Contributions I.S. performed the measurements. Both authors contributed to the analysis and writing the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.F. (jay@mail.huji.ac.il).

METHODS

Sample construction. Our experiments were conducted using poly(methylmethacrylate) (PMMA) blocks of dimensions 200 mm × 100 mm × 5.5 mm (top block) and 290 mm × 28 mm × 30 mm (bottom block) in the x , y and z direction, respectively (see Fig. 1a). Material shear, C_S , and longitudinal, C_L , wave speeds were obtained by measuring the time of flight of ultrasonic pulses, yielding $C_S = 1,345 \pm 10 \text{ m s}^{-1}$ and $C_L = 2,700 \pm 10 \text{ m s}^{-1}$. These provide $C_R = 1,255 \pm 10 \text{ m s}^{-1}$, where C_R is the Rayleigh wave speed. The values yield dynamic values for the Poisson ratio of $\nu \approx 1/3$ and Young's modulus of $E = 5.65 \text{ GPa}$. Note the value of E is significantly different from the static value of $E = 3 \text{ GPa}$. This difference is due to the viscoelastic behaviour of PMMA³¹. The contact face of the top block was diamond-machined to optical flatness. The bottom block's contact face was carefully polished first by 15- μm and later by smoother 9- μm diamond standard lapping film, yielding a 3- μm r.m.s. surface roughness.

Loading system. While the loading system was slightly modified and fitted for the purposes of the current research, the experimental system is described in detail elsewhere^{19,30}. The top block is clamped at its top edge, while the bottom block is mounted on a low-friction linear translational stage (Fig. 1a). External normal load, F_N , was fixed at the beginning of an experiment by pressing the blocks together in the y direction. Subsequently, external shear force, F_S , was quasi-statically (40 N s^{-1}) applied (starting from $F_S = 0$) to the bottom block in the negative x direction via a load cell of stiffness 10^7 N m^{-1} . Generally, this procedure results in an induced torque and consequently, strong gradients of normal stresses at the interface (see figure 1 in refs 19, 30). To partially compensate for this effect we varied the loading conditions by slightly tilting the top block before applying F_N .

Real contact area measurements. Changes in the real contact area along the entire interface were measured by an optical method based on total internal reflection. Basic principles are presented in detail elsewhere^{32,33}. A sheet of light, incident on the frictional interface at an angle well beyond the critical angle for total internal reflection, is reflected everywhere except at contact points (Extended Data Fig. 1a). This yields an instantaneous transmitted light intensity proportional to A over the entire ($x \times z$) $200 \times 5.5 \text{ mm}$ interface. The transmitted light is continuously imaged (at a spatial resolution of $1,280 \times 8$ pixels) at 580,000 frames per second using a high speed camera, Phantom v710. Data acquisition is continuous. The data are temporarily stored in a circular buffer large enough to acquire 7 ms of data both before and after each event. Storage of the data for each dynamic slip event is triggered by a sensitive acoustic sensor that is coupled to the top plate of the system.

The frictional interface is quasi-1D, as its width (z direction), 5.5 mm, is much smaller than any other dimensions of the block. The simultaneous measurements of $A(x, t)$ along the entire 1D interface is obtained by integration of the acquired images over the 8 pixels in the z direction.

We use a high power LED (CBT-90) as our source of light. This is in contrast to refs 32 and 33, where a laser sheet was used. One main advantage of a non-coherent light source is the absence of interference patterns and, therefore, a higher signal to noise ratio. Moreover, PMMA is a photo-elastic material, and non-polarized light neutralizes photoelastic effects that could affect the intensity of the transmitted light.

Strain measurements. We use miniature Vishay 015RJ rosette strain gauges for local measurements of a 2D strain tensor. 19 such strain gauges are mounted along and $\sim 3.5 \text{ mm}$ above the frictional interface, on one side of the upper block (Fig. 1a). All strain signals (57 channels) are amplified (gain ~ 400 , $\sim 1 \text{ MHz}$ bandwidth) and simultaneously acquired to 14 bit accuracy by an ACQ132 digitizer (D-tAcq Solutions Ltd) which simultaneously acquires all strain signals at a 1 MHz rate. Acquisition is synchronized with the contact area measurements.

Dimensions of the strain gauges are presented in Extended Data Fig. 1b. The active size of each strain component is $0.34 \times 0.38 \text{ mm}$. The measured strain components, $\varepsilon_1, \varepsilon_2, \varepsilon_3$, are later converted to $\varepsilon_{xx}, \varepsilon_{yy}$ by means of simple linear transformations:

$$\begin{aligned}\varepsilon_{yy} &= \varepsilon_1 \\ \varepsilon_{xy} &= 1/2(\varepsilon_3 - \varepsilon_2) \\ \varepsilon_{xx} &= \varepsilon_3 + \varepsilon_2 - \varepsilon_1\end{aligned}\quad (3)$$

Direct displacement, u_x , measurements. Two Philtec D20 optical displacement sensors were used to monitor the local motion in the x direction, u_x , at different locations along the top block (Extended Data Fig. 1c, for clarity only one is shown). The original operation method of the Philtec sensor is based on measuring the reflectance of the transmitted light from a mirror target moving in a direction parallel to the light beam. As we need to measure displacement, u_x , perpendicular to the sensor's axis (Extended Data Fig. 1c, right), we used the sensor in a different way—making use of the fact that the sensor is, in essence, a reflectance sensor.

Thus, if a larger area of the light spot is incident on the reflective target, a larger signal is measured. As the target moves (in the x direction), different fractions of the spot area are reflected and the measured signal varies with u_x (Extended Data Fig. 1b, left). This signal is calibrated, and a $\sim 0.1 \mu\text{m}$ resolution at an acquisition rate of 1 MHz is obtained. The sub-millimetre size of these sensors allows us to therefore obtain local measurements at extremely high precision and rapid rates.

Local material velocity, \dot{u}_x , measurements. Local material velocity, \dot{u}_x , is determined by the numerical time derivative of the displacement, u_x , measured by the displacement sensor. Measurements of u_x , however, require delicate mounting and calibration of the displacement sensor. It is, therefore, extremely difficult to mount several sensors along the interface. In what follows, a more easily accessible technique is presented.

If the rupture propagation velocity, C_f , varies adiabatically, we can assume $u_x(x, t) = u_x(x - C_f t)$. As $\varepsilon_{xx} = \frac{\partial u_x}{\partial x}$, we can obtain the local material velocity via $\frac{\partial u_x}{\partial t} = -C_f \varepsilon_{xx}$ (for convenience, in the text we denote $\frac{\partial u_x}{\partial t}$ by \dot{u}_x). Therefore, the measured ε_{xx} combined with measurement of C_f can be used to measure the material velocities at each strain gauge site. As we measure ε_{xx} in 19 locations along the interface and C_f is extracted from the instantaneous location of the rupture tip (measurements of $A(x, t)$), this technique enables a relatively dense measurement of \dot{u}_x along the interface without any additional preparations.

This technique was checked explicitly in several experiments, when displacement sensors were mounted slightly below one of the strain gauges and simultaneously measured. Extended Data Fig. 2 shows typical experiments where this technique is compared with direct displacement measurements. As the figure shows, the agreement between the two techniques is often within our measurement error.

Converting temporal to spatial measurements. The distance between adjacent strain gauges ($\sim 1 \text{ cm}$) is too large to provide good resolution of the spatial variation of the rapidly propagating fronts (Fig. 2b). We implement a high temporal resolution ($1 \mu\text{s}$) to overcome this difficulty. We assume that locally, the rupture is quasi-steady, $\varepsilon_{ij}(x, t) = \varepsilon_{ij}(x - C_f t)$. Therefore, temporal measurements of $\varepsilon_{ij}(x = x_0, t)$ (x_0 is the position of the strain gauge) and C_f , together with the transformation $t \rightarrow -tC_f$, allow us to dramatically increase our spatial resolution. Extended Data Fig. 2 demonstrates typical examples where this quasi-steady assumption was verified. We compensated for slow spatial variations of $C_f(x)$ by using $\varepsilon_{ij}(x, t) = \varepsilon_{ij}(x - \int C_f dt)$ instead of $\varepsilon_{ij}(x, t) = \varepsilon_{ij}(x - C_f t)$, where $C_f(t)$ was obtained from $A(x, t)$ measurements.

Comparing LEFM to measured strains. Strain gauge finite dimensions. Both the finite dimensions (Extended Data Fig. 1b) of the strain gauges and the temporal resolution of the acquisition system should be taken into account for proper interpretation of the measured strain signals of rapidly propagating ruptures. In order to accurately compare theory and measurements we convert the strain components $\varepsilon_{xy}, \varepsilon_{xx}, \varepsilon_{yy}$, predicted by LEFM, to $\varepsilon_1, \varepsilon_2, \varepsilon_3$ (for orientation of the measured strain components, see Extended Data Fig. 1b) by the inverse transformation of equation (3). After $\varepsilon_1, \varepsilon_2, \varepsilon_3$ were spatially averaged in accordance with both the geometry of the strain gauge components and the temporal differences due to the finite distances between the components, theoretical values corresponding to the measured ε_{ij} were obtained. These were the basis of the comparisons performed in Figs 2 and 3.

Boundary conditions. The universal LEFM solution is given¹⁰ for both plane stress ($\sigma_{zz} = 0$) and plane strain ($\varepsilon_{zz} = 0$) boundary conditions. Plane stress boundary conditions are considered to be relevant when a thin plate with traction-free faces is considered. This assumption should be violated close to the frictional interface where the material is pinned—and therefore not allowed to expand freely in z direction. Moreover, when short wavelength (compared to the width of the sample) signals are considered, plane strain boundary conditions are generally more appropriate. This might be relevant for rapid ruptures.

In our experiments we expect that some mixture of the two boundary conditions might take place. Extended Data Fig. 3 demonstrates that while both solutions have similar characteristic features some differences are apparent; measurements of ε_{xx} are slightly better described by the plane stress solution, whereas measurements of ε_{yy} are significantly better described by plane strain.

These considerations, however, are beyond the scope of the current work and in the main text we compare our measurements with the solution dictated by plane strain boundary conditions alone.

LEFM coupled to simple cohesive zone model. The universal square-root singularity, predicted by LEFM, must be regularized by dissipative processes taking place in the vicinity of the crack tip. Therefore, it should be interpreted as 'intermediate asymptotics'; the universal functional form is expected to describe the elastic fields at distances far from the interface weakening region, X_c , but still small compared to other dimensions of the system¹¹. This assumption is known as 'small scale yielding'. For high C_f , X_c is sufficiently small to have a clear separation of

scales. In the slow rupture regime the measurements were performed at a height $h = 3.5$ mm above the interface, which is a scale that is comparable to X_c (~ 2.5 mm at low velocities). The measured strains, however, are surprisingly well-described by the theory. This observation is perfectly compatible with simple cohesive zone models^{11,22}, as we demonstrate in Extended Data Fig. 4.

The effect of higher order terms. The complete LEFM stress field consists of a singular, $r^{-1/2}$, term together with additional (non-universal) terms of the form $r^{1/2}$, $r^{3/2}$, and so on, which vanish at the tip but may be significant at distances away from the crack tip.

Following the general solution procedure described in refs 10 or 22 we can derive the full angular dependence of these contributions to the stress field:

$$\sigma_{xy}(r, \theta) = C_n (2\pi r)^{\frac{n}{2}} \frac{1}{D(v)} \left(4\alpha_d \alpha_s \gamma_d^2 \cos\left(\frac{n\theta_d}{2}\right) - (1 + \alpha_s^2)^2 \gamma_s^2 \cos\left(\frac{n\theta_s}{2}\right) \right)$$

$$\sigma_{yy}(r, \theta) = -C_n (2\pi r)^{\frac{n}{2}} \frac{2\alpha_s (1 + \alpha_s^2)}{D(v)} \left(\gamma_d^2 \sin\left(\frac{n\theta_d}{2}\right) - \gamma_s^2 \sin\left(\frac{n\theta_s}{2}\right) \right)$$

where the functions $\alpha_s, \alpha_d, \gamma_s, \gamma_d, \theta_s, \theta_d, D$ are defined in ref. 10, $n = -1, 1, 3, 5, \dots$ (odd terms are considered here) and C_n are numerical constants. For $n = -1$ we reconstruct the universal form with $C_{-1} = K_{II}$. Generally, for $n > -1$ the constants C_n are not universal and cannot be easily determined.

We now demonstrate that these terms will not resolve the discrepancy, observed for high C_6 in $\Delta\epsilon_{xy}$ between the singular LEFM term and the measured signal, as presented in Fig. 3b, bottom. For simplicity, we consider the contribution of the terms $n = 1, 3$. We attempt to resolve the discrepancies in $\Delta\epsilon_{xy}$ in the range $0 < x - x_{tip}$ by varying C_1 and C_3 under the constraint that we can still reasonably describe $\Delta\epsilon_{xx}$ and $\Delta\epsilon_{yy}$. As Extended Data Fig. 5 demonstrates, these terms are insufficient to resolve the $\Delta\epsilon_{xy}$ discrepancies in the range $0 < x - x_{tip} < 25$ mm while still demanding consistency with the measured data for $|x - x_{tip}| > 25$ mm. Assuming that the next terms ($n > 3$) will dominate at even larger distance, we conclude that, within this framework, the observed discrepancy is not resolved.

Estimating cohesive zone properties. Our measured values of Γ and X_c can provide rough estimates of cohesive zone properties. Let us assume a slip weakening model^{11,12} in which the stress singularity predicted by LEFM is regularized by a peak shear stress, τ_p , that degrades to a residual strength, τ_r , over a slip distance, d_c . For simplicity, the weakening within the cohesive zone is assumed to be linear. It is then found that for slow ruptures, $X_c^0 = \frac{9\pi}{32} \times \frac{K_{II}^2}{(\tau_p - \tau_r)^2}$, where K_{II} is calculated from $\Gamma = [(1 - \nu^2)/E] K_{II}^2$

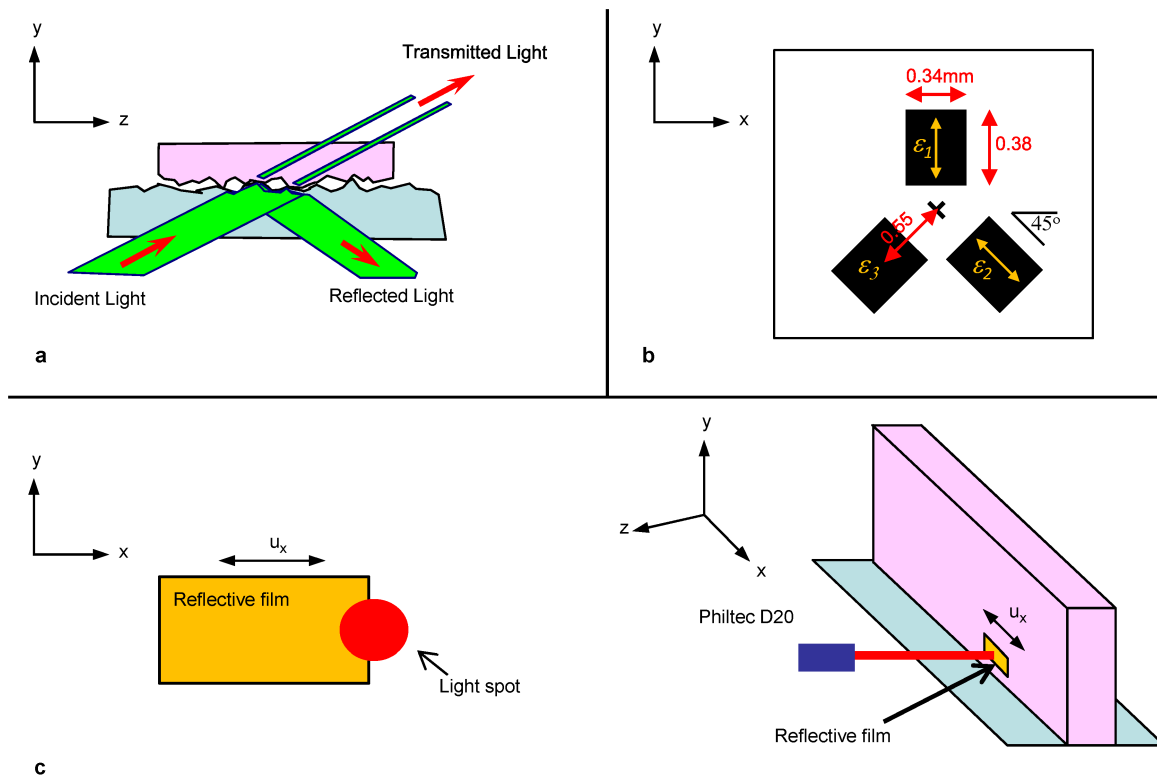
(ref. 11). Using the values Γ and X_c^0 from the text and the dynamic values of the Poisson ratio and Young's modulus, we find $\tau_p - \tau_r \approx 1.6 \pm 0.4$ MPa. For the conditions applied in the current experiments, $\tau_r \approx 3.7 \pm 0.4$ MPa. This leads to an estimation of $\tau_p \approx 5.3 \pm 0.8$ MPa. The linear slip weakening model also estimates that $\Gamma = 1/2(\tau_p - \tau_r)d_c$. This yields $d_c \approx 1.4$ μm , which is consistent with the asperity size of our rough interface.

Using this value of d_c , we can estimate, using simple scaling arguments^{27,34}, the maximal slip velocity $v_{\max} \approx C_f d_c / X_c$. Thus, the contraction of X_c in Fig. 4c indicates strong growth of v_{\max} , which was shown in recent numerical simulations³⁴. Using $\dot{u}_x = -C_f \dot{\epsilon}_{xx}$ we obtain, for example in Fig. 2b at $C_f = 0.96 C_R$, a maximal value of $\dot{u}_x \approx 0.3 \text{ m s}^{-1}$ measured 3.5 mm above the interface. At the interface, for the same C_f , the model then predicts 7 times larger velocity ($\dot{u}_x \approx 2 \text{ m s}^{-1}$). Following this procedure, our maximal measured velocity of $\sim 1 \text{ m s}^{-1}$ (Fig. 3a) corresponds to at least $\dot{u}_x \approx 10 \text{ m s}^{-1}$ at the interface.

This discussion suggests that caution should be used when interpreting measurements at a finite distance from the interface as representative of processes at the interface.

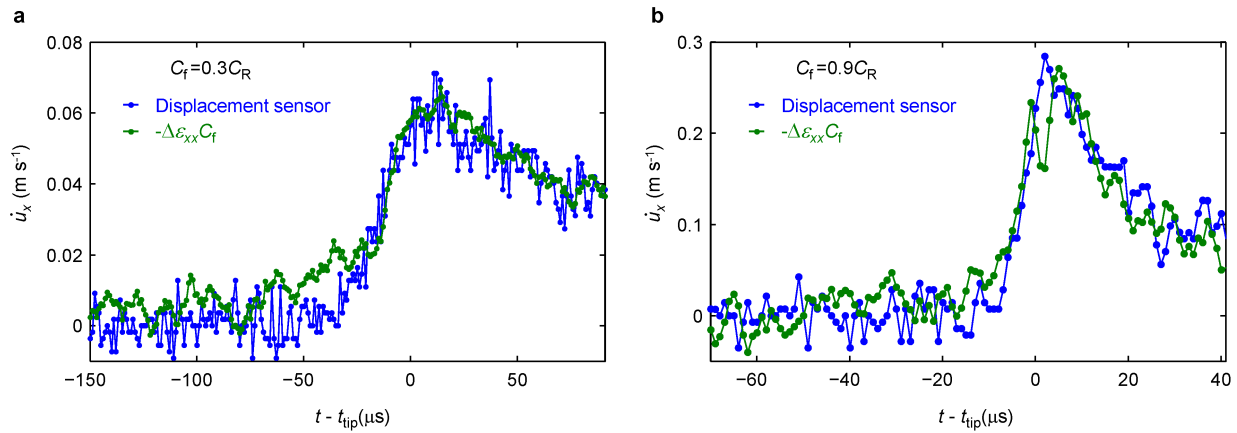
Considering the value of Γ . For our experimental conditions, A is approximately 0.5% of the nominal contact area⁴. For different applied normal stresses this value will change. In addition, A is typically reduced (Fig. 4a) by about 20% during the rupture process. As a result, the value of $\Gamma \approx 1.1 \text{ J m}^{-2}$ results from a reduction of approximately 0.1% of the nominal fracture area. Γ therefore corresponds to a bulk fracture energy (a 100% reduction of the nominal fracture energy) of about $\Gamma_{\text{Bulk}} \approx 1,100 \text{ J m}^{-2}$. This value is close to the measured value of Γ_{Bulk} for mode I fracture experiments in PMMA³⁵ and had been previously estimated for frictional dissipation using a thermal argument².

31. Read, B. E. & Duncan, J. C. Measurement of dynamic properties of polymeric glasses for different modes of deformation. *Polym. Test.* **2**, 135–150 (1981).
32. Rubinstein, S. M., Shay, M., Cohen, G. & Fineberg, J. Crack-like processes governing the onset of frictional slip. *Int. J. Fract.* **140**, 201–212 (2006).
33. Rubinstein, S. M., Cohen, G. & Fineberg, J. Visualizing stick-slip: experimental observations of processes governing the nucleation of frictional sliding. *J. Phys. D* **42**, 214016 (2009).
34. Gabriel, A. A., Ampuero, J. P., Dalguer, L. A. & Mai, P. M. Source properties of dynamic rupture pulses with off-fault plasticity. *J. Geophys. Res.* **118**, 4117–4126 (2013).
35. Sharon, E., Gross, S. P. & Fineberg, J. Energy dissipation in dynamic fracture. *Phys. Rev. Lett.* **76**, 2117–2120 (1996).



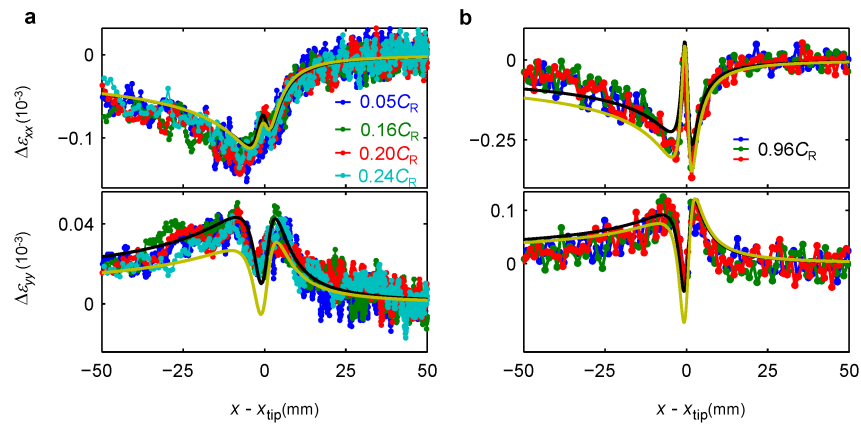
Extended Data Figure 1 | Experimental techniques. **a**, A method based on total internal reflection measures instantaneous changes in the real contact area, $A(x,t)$, along the entire interface. A sheet of light incident on the frictional interface is totally reflected everywhere except at contact points. Top block, pink; bottom block, blue. **b**, Geometry and dimensions (in mm) of a single rosette strain gauge. The black rectangles represent the active area of the measuring components, ϵ_1 , ϵ_2 and ϵ_3 . Yellow arrows represent the direction of

the measured strains. **c**, By placing a reflective displacement sensor at the edge of a reflecting film (left), the change in the reflected signal is a monotonic function of the displacement of the film. The reflecting area was calibrated to obtain high resolution measurements of the motion of the strip edge in the x direction (right). **a–c**, Measurements described above were acquired continuously. A sensitive acoustic sensor, mounted on the top block, triggered storage of all data bracketing every slip event.



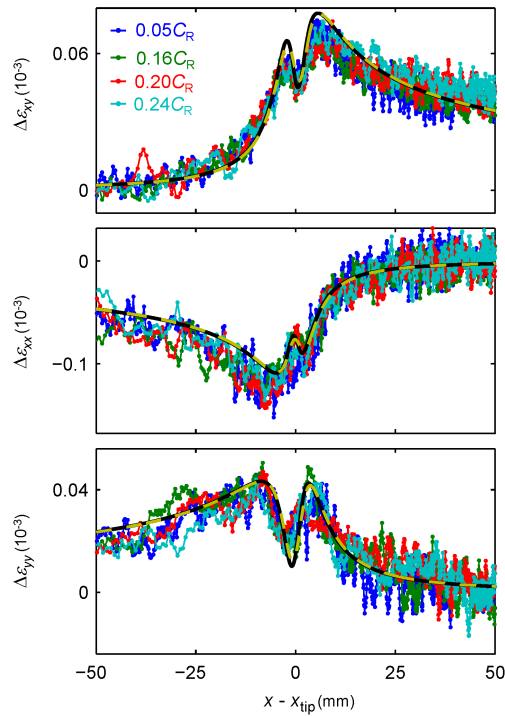
Extended Data Figure 2 | Comparing different techniques for measuring \dot{u}_x . In several experiments, direct displacement measurements were performed slightly below the strain gauge. The numerical time derivative of the corresponding signal is presented in blue. We reconstruct \dot{u}_x by an alternative method (green) that is based on measuring C_f , $\Delta \varepsilon_{xx}$ and the assumptions

presented in the text. **a, b,** Comparisons of \dot{u}_x by using both methods for two typical examples at $C_f = 0.3 C_R$ (**a**) and $C_f = 0.9 C_R$ (**b**) show good agreement between the measurements. x axes are the times relative to the time, t_{tip} , when the rupture fronts passed each measurement location.

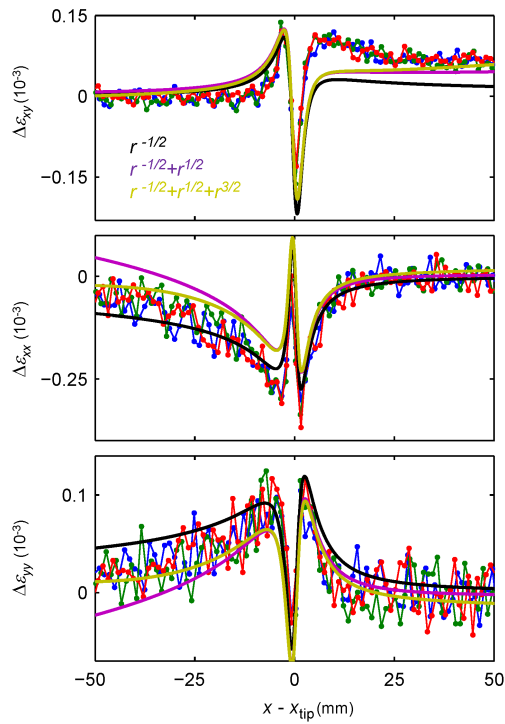


Extended Data Figure 3 | LEFM solution for plane strain and plane stress boundary conditions. **a, b,** Strain measurements of slow (**a**) and rapid (**b**) ruptures as presented in Fig. 2b. The top (bottom) panels are the strain components ϵ_{xx} (ϵ_{yy}) relative to their initial values immediately prior to rupture. The data presented are coloured as noted in the keys in the upper panel of **a** and

b. The solution for plane strain boundary conditions is shown in black. The solution for plane stress boundary conditions is shown in yellow. We see that plane strain conditions, as might be enforced by frictional pinning at the interface, provide a better fit to the data.



Extended Data Figure 4 | Comparing LEFM to linear cohesive zone model predictions. Strain measurements of a slow rupture as presented in Fig. 2a. Top, the strain component ε_{xy} , relative to the residual strain after the rupture passage. The centre (bottom) panels are the strain components ε_{xx} (ε_{yy}) relative to their initial values immediately prior to rupture. The data presented are coloured as noted in the key in the upper panel. Black solid line, the universal LEFM solution at 3.5 mm above the interface. Yellow dashed line, prediction of LEFM coupled to a linear slip weakening model^{11,22} for $X_c = 2.5$ mm and using the same value of I as used in the LEFM solution 3.5 mm above the interface. In the slow rupture regime, for $h/X_c \approx 1$ (h is distance of the strain gauge above the interface) linear slip weakening model is indistinguishable from the singular LEFM solution.



Extended Data Figure 5 | Contributions of higher-order terms to LEFM.

Strain measurements of a rapid rupture as presented in Fig. 2b. Top, the strain component ε_{xy} relative to the residual strain after the rupture passage. The centre (bottom) panels are the strain components ε_{xx} (ε_{yy}) relative to their initial values immediately prior to rupture. The data presented are for three different measurements in which $C_f = 0.96C_R$. The key in the upper panel denotes the terms that were considered for each coloured line. No value of the $r^{1/2}$ (C_1) and $r^{3/2}$ (C_3) coefficients could both resolve the discrepancy in $\Delta\varepsilon_{xy}$ and reasonably fit $\Delta\varepsilon_{yy}$ and $\Delta\varepsilon_{xx}$.

Tropical forcing of the recent rapid Arctic warming in northeastern Canada and Greenland

Qinghua Ding¹, John M. Wallace², David S. Battisti², Eric J. Steig¹, Ailie J. E. Gallant³, Hyung-Jin Kim⁴ & Lei Geng²

Rapid Arctic warming and sea-ice reduction in the Arctic Ocean are widely attributed to anthropogenic climate change^{1–3}. The Arctic warming exceeds the global average warming because of feedbacks that include sea-ice reduction^{4,5} and other dynamical and radiative feedbacks^{6–13}. We find that the most prominent annual mean surface and tropospheric warming in the Arctic since 1979 has occurred in northeastern Canada and Greenland. In this region, much of the year-to-year temperature variability is associated with the leading mode of large-scale circulation variability in the North Atlantic, namely, the North Atlantic Oscillation^{14,15}. Here we show that the recent warming in this region is strongly associated with a negative trend in the North Atlantic Oscillation, which is a response to anomalous Rossby wave-train activity originating in the tropical Pacific. Atmospheric model experiments forced by prescribed tropical sea surface temperatures simulate the observed circulation changes and associated tropospheric and surface warming over northeastern Canada and Greenland. Experiments from the Coupled Model Intercomparison Project Phase 5 (ref. 16) models with prescribed anthropogenic forcing show no similar circulation changes related to the North Atlantic Oscillation or associated tropospheric warming. This suggests that a substantial portion of recent warming in the northeastern Canada and Greenland sector of the Arctic arises from unforced natural variability.

The recent Fifth Assessment Report of the Intergovernmental Panel on Climate Change¹ anticipates the continuing retreat of sea ice and warming accompanying future anthropogenic emissions of greenhouse gases and aerosols. However, natural variability, such as that associated with the Atlantic Multi-decadal Oscillation⁶, has been suggested to be an important driver of climate variations in the Arctic region and responsible for a portion of the recent warming trend. Some recent results also indicate that sea surface temperature (SST) changes outside the Arctic have played a role in forcing the recent tropospheric warming in the Arctic¹⁷.

In this study, we use observational analyses and modelling to explore the relative contributions of anthropogenic forcing and natural variability to recent warming trends in the Arctic. We identify a specific tropical SST pattern that has contributed significantly to recent Arctic warming by changing the circulation over northern high latitudes.

We use post-1979 observations only, because the analyses of geopotential height (which approximates the actual height of a pressure surface above mean sea level) and other variables over the Northern Hemisphere polar region are more reliable during the modern satellite era¹⁸. The most reliable reanalysis data sets (ERA-interim¹⁹ and MERRA²⁰) and station data^{21,22} agree that the most pronounced annual mean surface temperature warming in the Arctic since 1979 has occurred over northeastern Canada, Greenland and the north of Siberia (Fig. 1a and Extended Data Fig. 1). Surface temperatures along the Siberian coast are highly negatively correlated ($r \approx -0.7$ to -0.9) with *in situ* sea-ice variability where the observed surface warming is greatest and can be related directly to the significant sea-ice reduction. The troposphere has also experienced an annual averaged warming in the Arctic. Tropospheric warming is most

pronounced over northeastern Canada and Greenland, where it extends through the depth of the troposphere and has occurred in all seasons, with the largest magnitude and extent during boreal winter (December–February) (Fig. 1b, Extended Data Figs 2 and 3). The magnitude of the surface and tropospheric warming in the northeastern Canadian–Greenland sector of the Arctic is nearly twice as large as the Arctic-mean warming.

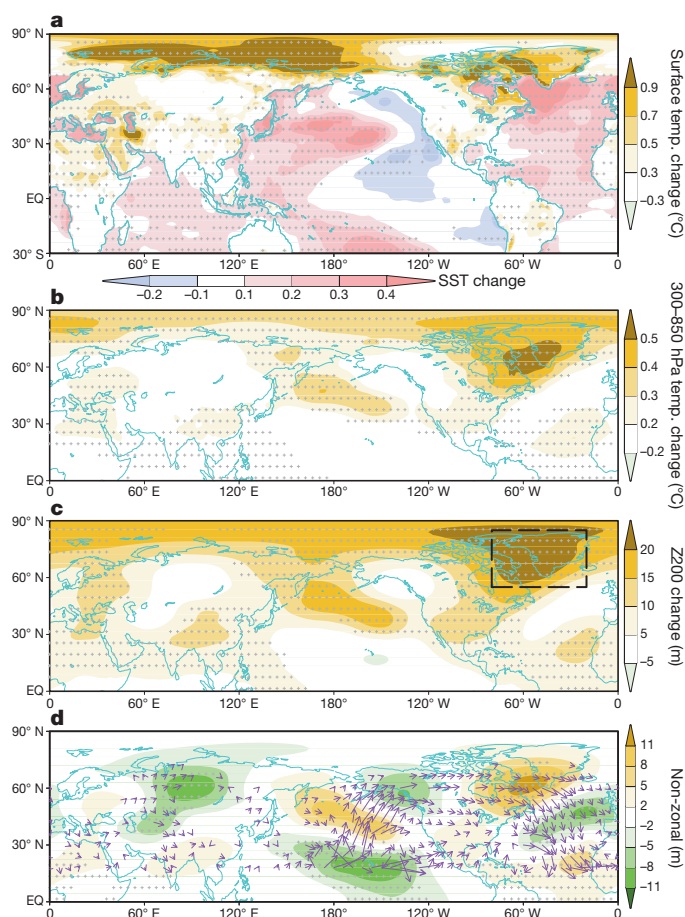


Figure 1 | Observed trend pattern of annual mean field for 1979–2012. Linear trend (per decade) of annual mean surface temperature (a), 300–850 hPa temperature (b), 200-hPa geopotential height (Z200; c) and the non-zonal component of 200-hPa geopotential height (d). In a, surface temperature is shown over land or ice; SST is shown over ocean. In d, purple vectors (units: $10^6 \text{ Pa m}^2 \text{ s}^{-2}$, vectors less than $10^5 \text{ Pa m}^2 \text{ s}^{-2}$ are omitted) denote the wave activity flux associated with the eddy Z200 trend pattern. Grid points with trends that are statistically significant at the 99% confidence level are denoted by small crosses. The box in c indicates the domain over which data are averaged in Extended Data Fig. 6. EQ, Equator.

¹Department of Earth and Space Sciences and Quaternary Research Center, University of Washington, Seattle, Washington 98195, USA. ²Department of Atmospheric Sciences, University of Washington, Seattle, Washington 98195, USA. ³School of Geography and Environmental Science, Monash University, Victoria 3800, Australia. ⁴Climate Research Department, APEC Climate Center, 12 Centum 7-ro, Haerundae-gu, Busan 612-020, South Korea.

The spatially varying character of warming trends over the Arctic region suggests a role for atmospheric circulation changes. Annual mean 200-hPa geopotential heights (Z200) in the upper troposphere have increased almost everywhere in the Northern Hemisphere polar region since 1979, with the greatest increases occurring over northeastern Canada and Greenland, coinciding with the greatest surface and tropospheric warming (Fig. 1c and Extended Data Fig. 4). Consistent with the local surface temperature trend, the positive trend in Z200 over northeastern Canada and Greenland has occurred in all seasons, and is largest during winter (Extended Data Fig. 5). The correlation (r) between 34-year annual mean Z200 and surface temperature in this region is 0.9, and it is 0.8 for the detrended component (Extended Data Fig. 6a). Both the Z200 and surface temperature time series exhibit decadal-scale variability superimposed on the long-term trend. In all seasons, the seasonal and decadal anomalies in Z200 are highly correlated with the surface temperature anomalies (Extended Data Fig. 6b–e).

Because the upper-level atmospheric circulation is relatively insensitive to anomalous surface warming in the polar regions^{7,17}, it seems unlikely that the decadal-scale variations in Z200 in the 34-year record could be forced by local variations in surface temperature. Theory and a myriad of modelling studies suggest that causality operates in the opposite direction: that is, that the changes in surface and tropospheric temperature are an adiabatic response to the changes in the upper-level circulation, which may have been remotely forced.

The circulation trend is marked by rising geopotential heights over Greenland and near-steady heights along $\sim 50^\circ$ N over the North Atlantic (Fig. 1c). The pattern can be characterized as a uniform increase in geopotential heights across the Arctic, combined with the negative polarity of the North Atlantic Oscillation (NAO). The NAO is the leading mode of circulation variability and the main driver of temperature variability and extreme winter weather events in the North Atlantic^{14,15}. A year-round NAO index, given by the leading principal component of low-level geopotential height anomalies in the North Atlantic¹⁴, exhibits a strong ($r = 0.7$) correlation with surface temperature and Z200 anomalies in northeastern Canada and Greenland (Extended Data Fig. 6a). The past three decades have been marked by a negative trend in the NAO index²³.

Further insight into the causes of the remarkable warming trend over Greenland and northeastern Canada is gained by examining the trends

in the eddy component of Z200, obtained by removing the zonal average at each latitude from the trends (Fig. 1c). The results (Fig. 1d) strongly suggest that the 200-hPa geopotential height increases over northeastern Canada and Greenland are associated with a Rossby wave train that originates in the tropical Pacific. The corresponding flux of wave activity exhibits a distinctive arc-shaped trajectory, extending from the central tropical Pacific towards the Arctic and returning to the tropics via the Atlantic sector. Coincident with the pattern of trends in the warming and anomalous circulation, SSTs (and rainfall) have increased significantly at the 99% confidence level throughout the tropics except for the central and eastern Pacific where they have remained nearly constant or even decreased slightly, as shown in Fig. 1a (Extended Data Fig. 7a). Motivated by two earlier studies^{24,25}, we hypothesize that this pattern of SST change in the tropics has played a key role in forcing the extratropical wave train that has both warmed northeastern Canada and Greenland and induced the negative trend in the NAO in recent decades.

Our hypothesis that the trend towards higher geopotential heights over northeastern Canada and Greenland has occurred in response to trends in tropical SST is supported by maximum covariance analysis (MCA)²⁶. Figure 2 shows the leading modes of co-variability between annual mean Northern Hemisphere Z200, and tropical SST for the period 1979–2012. The leading MCA mode explains 68% of the squared covariance between the hemispheric Z200 and tropical SST fields, and captures the typical El Niño/Southern Oscillation (ENSO) signature in SST and the related atmospheric teleconnection pattern over the Pacific and North American sectors, with their familiar sequence of year-to-year variability. The patterns associated with the second mode resemble the trend in tropical SST and the wave-train pattern of the Z200 trend (Fig. 1a, c). The second MCA mode explains 28% of the squared covariance. The time series associated with this mode are strongly correlated ($r = 0.91$) and both exhibit a pronounced upward trend over the 34-year record. The time series of the SST pattern in mode 2 is highly correlated with the surface temperature in northeastern Canada and Greenland and with the year-round NAO index (Extended Data Fig. 8). This mode is not merely a reflection of the trends in the two fields, because similar MCA results are obtained when the data are detrended before conducting the MCA calculation (Extended Data Fig. 9). Thus, the wave train that links the positive temperature trends in northeastern Canada and Greenland to the tropical SST anomalies characterizes both

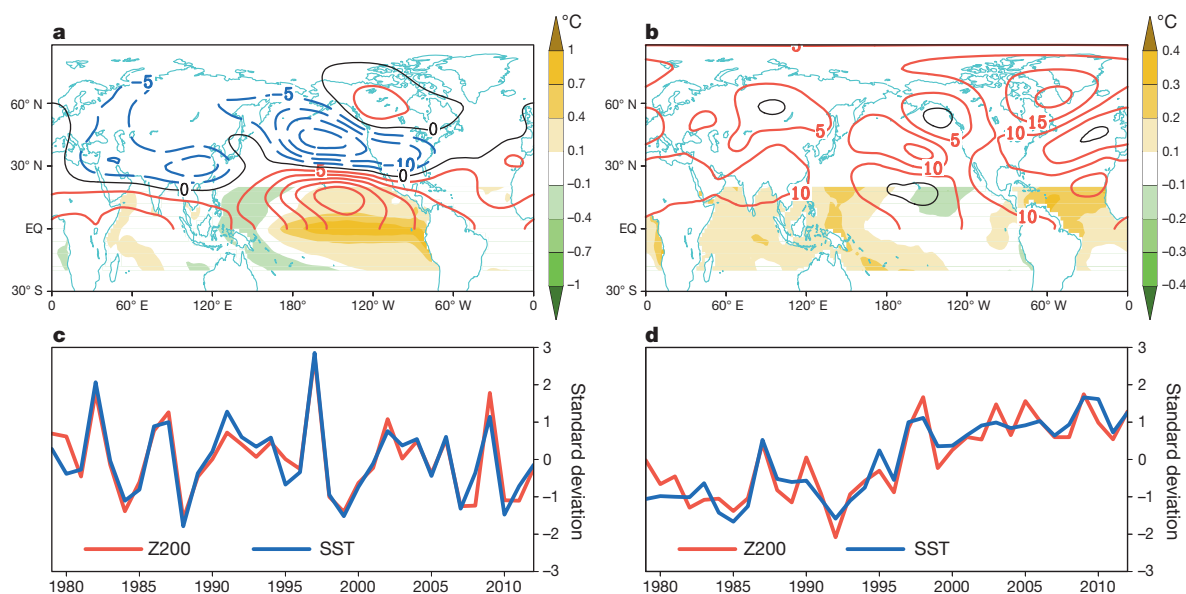


Figure 2 | Coupled patterns between annual mean tropical SST and Northern Hemisphere circulation for 1979–2012. a–d, Results of maximum covariance analysis for annual mean 1979–2012 Northern Hemisphere (0–88.5° N) 200-hPa geopotential height (Z200) and tropical (20° S to 20° N) SST. Shown in a are the patterns of Z200 (contours, interval 10 m) and tropical

SST (shading; colour scale) that accompany the first mode. c, Time series of the Z200 (red) and SST (blue) patterns shown in a. Panels b and d are the same as panels a and c, but for the second mode. Amplitudes in a and b are scaled by one standard deviation of the corresponding time series in c and d.

interannual and interdecadal timescales. Compared to the circulation anomalies associated with typical ENSO, the circulation anomalies associated with the mode 2 pattern of SST anomalies exhibits a wave train that extends farther poleward and includes circulation anomalies over the North Atlantic that are very similar to those associated with the NAO. Thus, the recent 30-year trends in tropical SST—distinct from ENSO—may be partially responsible for the negative trend in the conventional NAO index.

To further examine the relationship between tropical SST forcing and circulation variability in northeastern Canada and Greenland, we use an atmospheric general circulation model (ECHAM 4.6 (ref. 27)) to perform an ensemble of integrations, each of which is forced by the observed tropical SST from 1979–2012; in the extratropics, the atmosphere is coupled to a slab ocean with a thermodynamic sea-ice component. Ten experiments were run, each with different initial atmospheric conditions, and averaging over the ten experiments is used to filter out intrinsic atmospheric variability. The ensemble-averaged change in the high-latitude circulation is, by construction, due to the remote tropical forcing and to the local amplification associated with sea-ice change. The ensemble-averaged trends in surface temperature, tropospheric temperature and Z200 in the Northern Hemisphere are shown in Fig. 3; in the western hemisphere they closely resemble the observed trends shown in Figs 1a–c, and are consistent with the trend patterns associated with

the mode 2 in the MCA analysis. Rising 200-hPa heights and widespread warming over northeastern Canada and Greenland are an integral part of the response to the imposed tropical SST anomalies.

As in the observations, the simulated circulation anomaly over the North Atlantic is similar to that observed during the negative phase of the NAO. As in the observations, the warming trends and atmospheric circulation anomalies are largest in winter. The simulated annual mean Z200, surface and tropospheric temperature trends over northeastern Canada and Greenland are about 10 m per decade, 0.4°C per decade and 0.3°C per decade respectively, roughly half the magnitude of the values in the ERA-interim reanalysis¹⁹. We note that the model, forced by the observed tropical SST change, also reproduces the observed rainfall trend in the tropical Pacific (Extended Data Fig. 7b). The simulated wave-train path over the North Pacific and North Atlantic differs in detail from that observed (compare Figs 3d and 1d), possibly owing to subtle differences between the real and simulated climatological mean zonal flow. Nonetheless, the model results strongly suggest that a Rossby wave train associated with the observed tropical SST trend is responsible for the observed negative trend of the NAO and accounts for both the regional surface and tropospheric warming in northeastern Canada and Greenland.

Last, we explore the contribution of external forcing to the recent regional warming in northeastern Canada and Greenland. We take the ensemble-average response of the 40 climate models that were forced by historical forcing and archived in the CMIP5 database historical experiment¹⁶. We note that in these experiments, “historical forcing” includes both anthropogenic and natural forcing.

The trends in the ensemble-average surface and tropospheric temperature from the historical forcing simulations are shown in Fig. 4a and b. Comparing Fig. 1a to Fig. 4a, we see that the ensemble average of

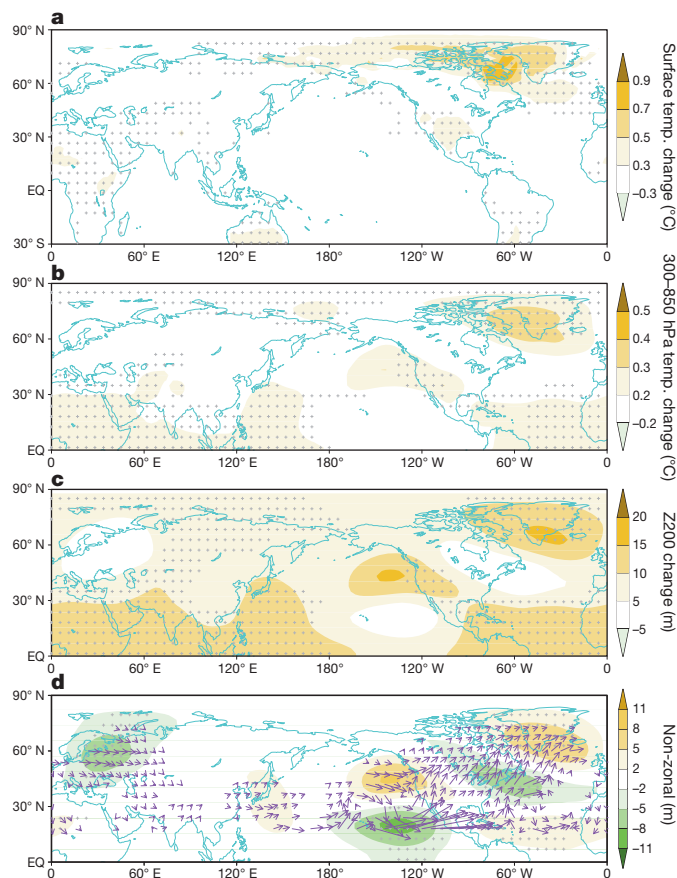


Figure 3 | Simulated 1979–2012 trend patterns in annual mean fields in ECHAM. Linear trend (per decade) of annual mean surface temperature (a), 300–850 hPa temperature (b), Northern Hemisphere 200-hPa geopotential height (Z200; c) and the non-zonal component of 200-hPa geopotential height from 34-year simulation of ECHAM run forced by observed SST (1979–2012) in the tropics (30°N to 30°S) (d). In the extratropics, the atmosphere is coupled to a slab ocean model. In d, purple vectors (units: $2.5 \times 10^5 \text{ Pa m}^2 \text{ s}^{-2}$, vectors less than $2.5 \times 10^4 \text{ Pa m}^2 \text{ s}^{-2}$ are omitted) denote the wave activity flux associated with the eddy Z200 trend pattern. Grid points with trends that are statistically significant at the 99% confidence level are denoted by small crosses.

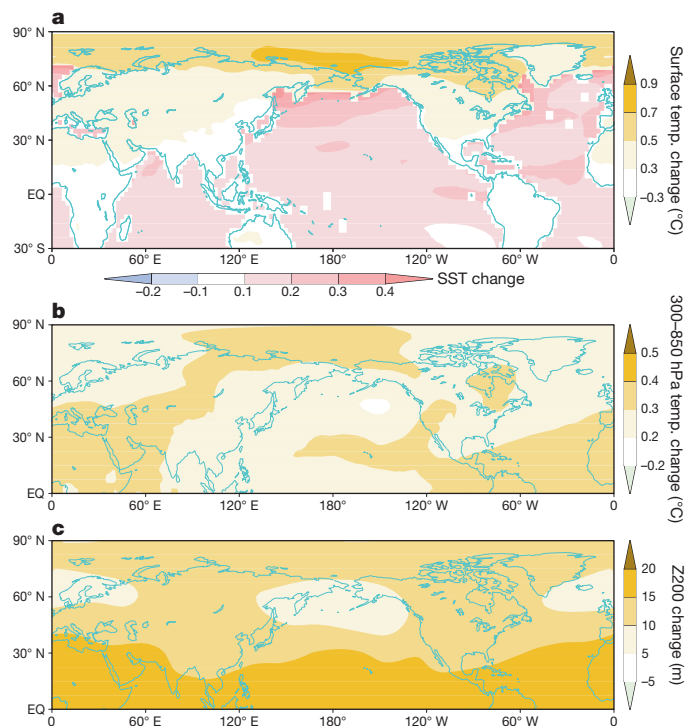


Figure 4 | Simulated 1979–2004 trend patterns in annual mean fields in CMIP5. Linear trend (per decade) of surface temperature (a), 300–850 hPa temperature (b) and 200-hPa geopotential height (Z200; c) for the period 1979–2004 due to anthropogenic and natural external forcing. Here we show the average of an ensemble of experiments using 40 climate models (Extended Data Table 1) that are archived in the CMIP5 database¹⁶ and used in the Fifth Assessment Report of the Intergovernmental Panel on Climate Change¹. The trends shown in the plot are significant at above the 99% confidence level at all grid points.

the historical simulations using the CMIP5 models shows a general warming throughout the globe—more so in the polar regions than in the tropics—with a maximum in the western Arctic Ocean due to reduced sea ice. The forced pattern of warming does not capture the cooling in the eastern tropical Pacific or the maximum warming over northeastern Canada and Greenland; in these regions, the CMIP5 models capture only about 40–50% of the observed warming in the surface and troposphere.

From a comparison of Figs 1c and 4c it is evident that the ensemble average of the historical simulations exhibits much larger Z200 rises in the tropics than observed, and slightly smaller increases in the polar region than observed. The most salient circulation feature in the observations is the pronounced localized rise in upper-tropospheric geopotential height and surface and tropospheric temperature over northeastern Canada and Greenland (Fig. 1). These features are not reproduced in the ensemble average of the historical simulations using the CMIP5 models.

The results from the ECHAM experiments forced by observed SST, together with the ensemble average of the historical simulations of the CMIP5 models, suggest that about half of the observed warming in northeastern Canada and Greenland is due to a uniform warming of the entire Arctic (directly attributable to anthropogenic forcing¹); the other half is due to a wave train of circulation anomalies that extends from the central tropical Pacific to Canada, and which is forced by trends in the tropical Pacific SST. That the historically forced CMIP5 model average does not capture the observed tropical SST pattern further suggests that around half of the recent warming over Canada and Greenland is associated with natural variability that is intrinsic to the coupled atmosphere–ocean system. We note, however, that the observational records in the tropics are temporally and spatially incomplete, and that the simulated response of tropical SSTs to anthropogenic forcing is strongly model-dependent²⁸. Hence we cannot rule out the possibility that the pattern of the observed tropical Pacific SST trend may also be associated with external forcing.

METHODS SUMMARY

We used atmospheric circulation and temperature data from the 1979–2013 ERA-interim reanalysis¹⁹, SST data from ERSSTv3 (ref. 29), and sea-ice data from HADISST³⁰. We used output from 40 climate models archived in the CMIP5 database¹⁶ to estimate the climate response to external forcing in the historical record. MCA²⁶ captured the dominant coupled modes between the Northern Hemisphere Z200 (0–88.5° N) change and tropical SST (20° N to 20° S). We used the ECHAM4.6 (ref. 27) atmospheric general circulation model to simulate the circulation and temperature response forced by the observed tropical (30° N to 30° S) SST trends; in the extratropics, the ECHAM4.6 is coupled to a slab ocean with a simple, thermodynamic-only sea ice model. We focus on the annual average in this study because the recent temperature and circulation changes over the region of interest are similar across all seasons. To construct an annual mean, we averaged the 12 months from June through to the end of the following May. Note that the year label refers to June (for example, the annual mean of 2012 is the average from June 2012 through to the end of May 2013). The results presented in this study are not sensitive to how the annual average is constructed. It should be noted that because the historical record of reliable data for the Northern Hemisphere polar region is short, the term ‘trend’ used in the main text refers to any low-frequency variability distinct from the inter-annual variability; it does not necessarily imply a long-term trend, but may simply reflect decadal variability.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 19 October 2013; accepted 12 March 2014.

1. Kirtman, B. *et al.* in *Climate Change 2013: The Physical Science Basis* (eds Stocker, T. F. *et al.*) 953–1028 (Cambridge Univ. Press, 2013).
2. Serreze, M. C. & Barry, R. G. Processes and impacts of Arctic amplification: a research synthesis. *Global Planet. Change* **77**, 85–96 (2011).
3. Polyakov, I., Walsh, J. E. & Kwok, R. Recent changes of Arctic multiyear sea ice coverage and the likely causes. *Bull. Am. Meteorol. Soc.* **93**, 145–151 (2012).

4. Serreze, M. C., Barrett, A. P., Stroeve, J. C., Kindig, D. M. & Holland, M. M. The emergence of surface-based Arctic amplification. *Cryosphere* **3**, 11–19 (2009).
5. Screen, J. A. & Simmonds, I. The central role of diminishing sea ice in recent Arctic temperature amplification. *Nature* **464**, 1334–1337 (2010).
6. Chylek, P., Folland, C. K., Lesins, G., Dubey, M. K. & Wang, M. Y. Arctic air temperature change amplification and the Atlantic multidecadal oscillation. *Geophys. Res. Lett.* **36**, L14801 (2009).
7. Graversen, R. G., Mauritsen, T., Tjernstrom, M., Kallen, E. & Svensson, G. Vertical structure of recent Arctic warming. *Nature* **451**, 53–56 (2008).
8. Lee, S. Testing of the tropically excited Arctic warming (TEAM) mechanism with traditional El Niño and La Niña. *J. Clim.* **25**, 4015–4022 (2012).
9. Francis, J. A. & Hunter, E. New insight into the disappearing Arctic sea ice. *Eos* **87**, 509–511 (2006).
10. Abbot, D. S., Walker, C. & Tziperman, E. Can a convective cloud feedback help eliminate winter sea ice at high CO₂ concentration? *J. Clim.* **22**, 5719–5731 (2009).
11. Hansen, J. & Nazarenko, L. Soot climate forcing via snow and ice albedo. *Proc. Natl Acad. Sci. USA* **101**, 423–428 (2004).
12. Shindell, D. & Faluvegi, G. Climate response to regional radiative forcing during the twentieth century. *Nature Geosci.* **2**, 294–300 (2009).
13. Yoo, C., Feldstein, S. & Lee, S. Impact of the Madden-Julian Oscillation trend on the polar amplification of surface air temperature during 1979–2008 boreal winter. *Geophys. Res. Lett.* **38**, L24804 <http://dx.doi.org/10.1029/2011GL049881> (2011).
14. Barnston, A. G. & Livezey, R. E. Classification, seasonality and persistence of low-frequency atmospheric circulation patterns. *Mon. Weath. Rev.* **115**, 1083–1126 (1987).
15. Hurrell, J. W. Decadal trends in the North Atlantic Oscillation: regional temperatures and precipitation. *Science* **269**, 676–679 (1995).
16. Taylor, K. E., Stouffer, R. J. & Meehl, G. A. An overview of the CMIP5 and the experimental design. *Bull. Am. Meteorol. Soc.* **93**, 485–498 (2012).
17. Screen, J. A., Deser, C. & Simmonds, I. Local and remote controls on observed Arctic warming. *Geophys. Res. Lett.* **39**, L10709 (2012).
18. Bromwich, D. H., Fogt, R. L., Hodges, K. I. & Walsh, J. E. A tropospheric assessment of the ERA-40, NCEP, and JRA-25 global reanalyses in the polar regions. *J. Geophys. Res.* **112**, D10111 (2007).
19. Dee, D. P. *et al.* The ERA-Interim reanalysis: configuration and performance of the data assimilation system. *Q. J. R. Meteorol. Soc.* **137**, 553–597 (2011).
20. Rienecker, M. M. *et al.* MERRA: NASA's modern-era retrospective analysis for research and applications. *J. Clim.* **24**, 3624–3648 (2011).
21. Willmott, C. J. & Matsuura, K. Smart interpolation of annually averaged air temperature in the United States. *J. Appl. Meteorol.* **34**, 2577–2586 (1995).
22. Hansen, J., Ruedy, R., Sato, M. & Lo, K. Global surface temperature change. *Rev. Geophys.* **48**, RG4004 (2010).
23. Hurrell, J. W. & Deser, C. North Atlantic climate variability: the role of the North Atlantic Oscillation. *J. Mar. Syst.* **78**, 28–41 (2009).
24. Hoerling, M. P., Hurrell, J. W. & Xu, T. Y. Tropical origins for recent North Atlantic climate change. *Science* **292**, 90–92 (2001).
25. Graf, H.-F. & Zanchettin, D. Central Pacific El Niño, the “subtropical bridge,” and Eurasian climate. *J. Geophys. Res.* **117**, D01102 (2012).
26. Wallace, J. M., Smith, C. & Bretherton, C. S. Singular value decomposition of wintertime sea surface temperature and 500-mb height anomalies. *J. Clim.* **5**, 561–576 (1992).
27. Roeckner, E. *et al.* *The Atmospheric General Circulation Model ECHAM-4: Model Description and Simulation of Present-Day Climate* (Max Planck Institut für Meteorologie Report 218, 1996).
28. Guilyardi, E. *et al.* New strategies for evaluating ENSO processes in climate models. *Bull. Amer. Meteorol. Soc.* **93**, 235–238 (2012).
29. Smith, T. M., Reynolds, R. W., Peterson, T. C. & Lawrimore, J. Improvements to NOAA's historical merged land-ocean surface temperature analysis (1880–2006). *J. Clim.* **21**, 2283–2296 (2008).
30. Rayner, N. A. *et al.* (2003) Global analyses of sea surface temperature, sea ice, and night marine air temperature since the late nineteenth century. *J. Geophys. Res.* **108**(D14), 4407, <http://dx.doi.org/10.1029/2002JD002670> (2003).

Acknowledgements We thank the Max Planck Institute for Meteorology model developer for making ECHAM4.6 available and C. Bitz, Q. Fu, D. L. Hartmann, D. Frierson and W.-J. Li for discussion. This work was supported by the US National Science Foundation (OPP 1043092 and ATM 1122989). Q.D. acknowledges support from the University of Washington's Quaternary Research Center and the National Basic Research Program of China (973 Program-2013CB430203). H.-J.K. acknowledges support from the APEC Climate Center.

Author Contributions Q.D. made the calculations, implemented the general circulation model experiments, created the figures, and led the writing of the paper. All authors contributed to the experimental design, and to writing the paper.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to Q.D. (qinghua@uw.edu)

METHODS

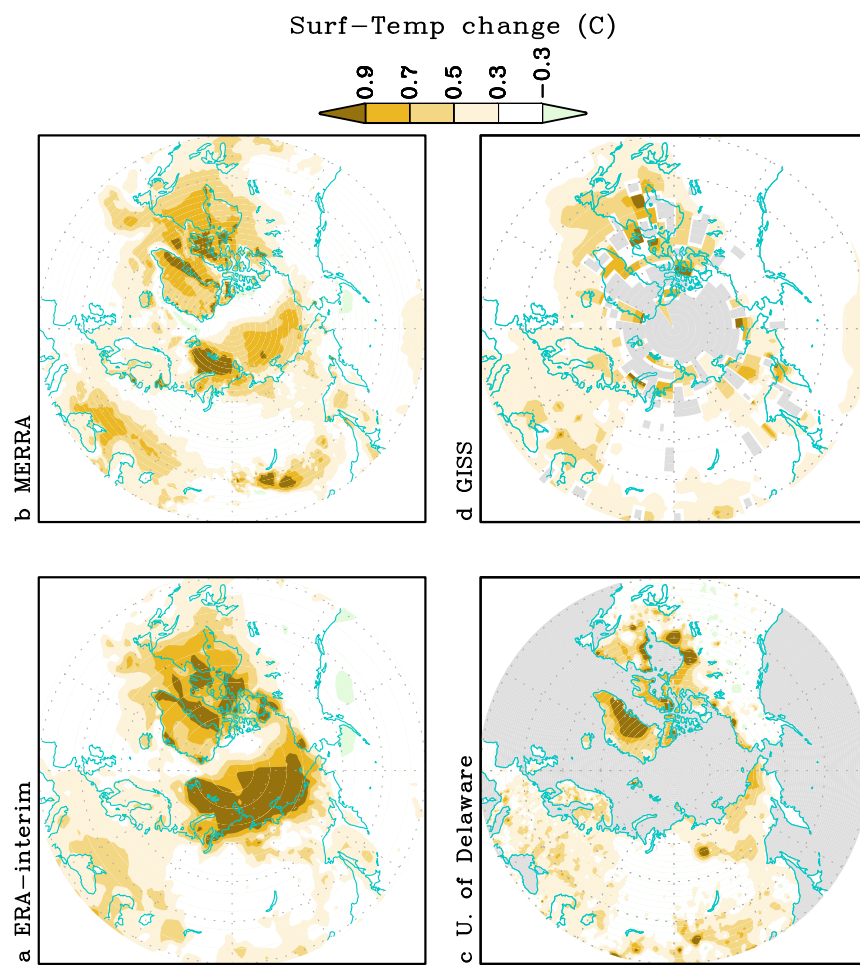
Data. Atmospheric circulation and temperature data are from the 1979–2013 ERA-Interim reanalysis¹⁹ and the Modern Era Retrospective-Analysis for Research and Applications (MERRA)²⁰ reanalysis. SST data are from ERSSTv3 (ref. 29). Sea-ice data are from HADISST³⁰. Rainfall data are from GPCP³¹. We also use land-only station surface and near-surface temperature data from University of Delaware²¹ and NASA GISS Surface Temperature (GISTEMP)²². We used output from the historical simulations of 40 climate models (Extended Data Table 1) that are archived in the CMIP5 (ref. 16) database.

MCA. We use maximum covariance analysis (MCA²⁶) to capture the dominant coupled modes between the Northern Hemisphere Z200 (0–88.5° N) change and tropical SST (20°N to 20°S). MCA organizes the variance in two data sets into orthogonal patterns that maximize the covariance in the data.

Wave flux analysis. Wave activity analysis³² is used to reveal stationary Rossby wave energy propagation. The wave flux vector provides direct information on the flux of wave activity, which is parallel to the group velocity of quasi-stationary Rossby waves. This diagnostic tool is well suited for detection of propagating large-scale quasi-stationary Rossby waves.

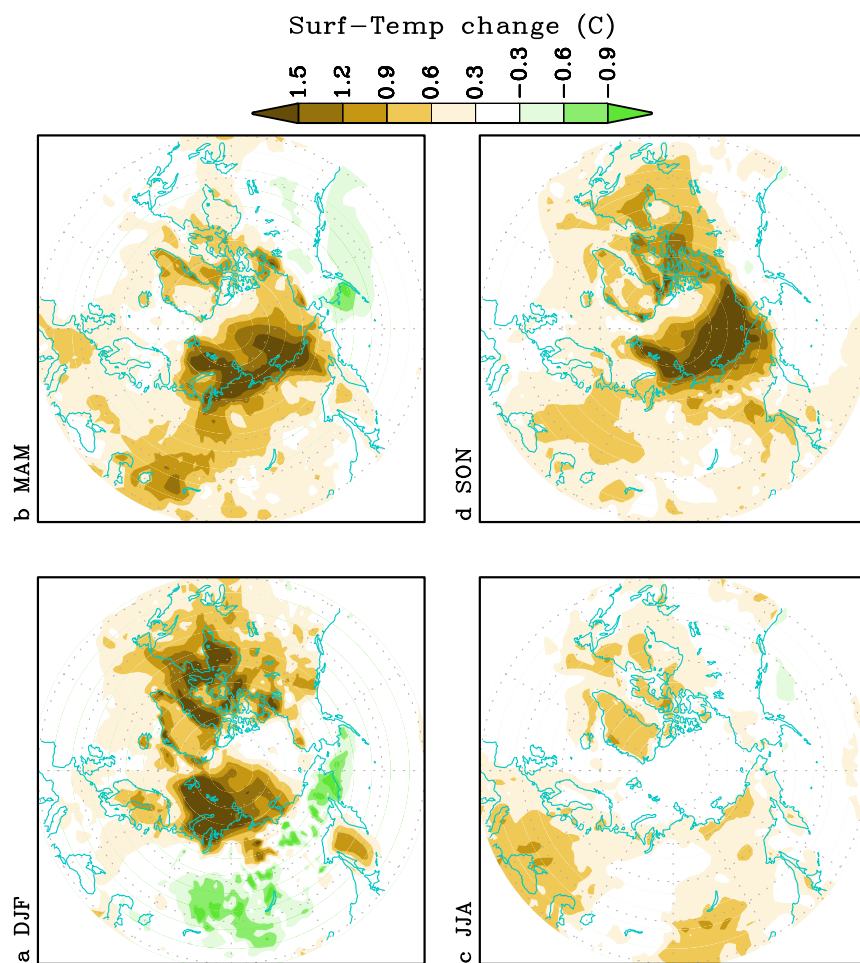
Climate model. The general circulation model used to perform the experiments in this study is the ECHAM4.6 atmospheric general circulation model²⁷, at horizontal resolution of T42 ($\sim 2.8^\circ$ latitude \times 2.8° longitude) with 19 vertical levels. We coupled the ECHAM4.6 to a slab ocean, whereby the ocean is represented as boxes of water of uniform specified depth (50 m). A simple, thermodynamic-only sea-ice model is included when the ocean temperature reaches the freezing point. The ocean temperature or sea-ice condition at each grid point is affected only by heat exchange across the air–sea interface; there is no direct communication between adjacent ocean grid points, nor is there any representation of the deep ocean. In addition, a cyclostationary climatological heat flux is added to the ocean temperature tendency equation in order to maintain a seasonal cycle of ocean temperature and sea ice condition that is as close to that observed.

31. Huffman, G. J., Adler, R. F., Bolvin, D. T. & Gu, G. Improving the global precipitation record: GPCP version 2.1. *Geophys. Res. Lett.* **36**, L17808 (2009).
32. Plumb, R. A. On the three-dimensional propagation of stationary waves. *J. Atmos. Sci.* **42**, 217–229 (1985).



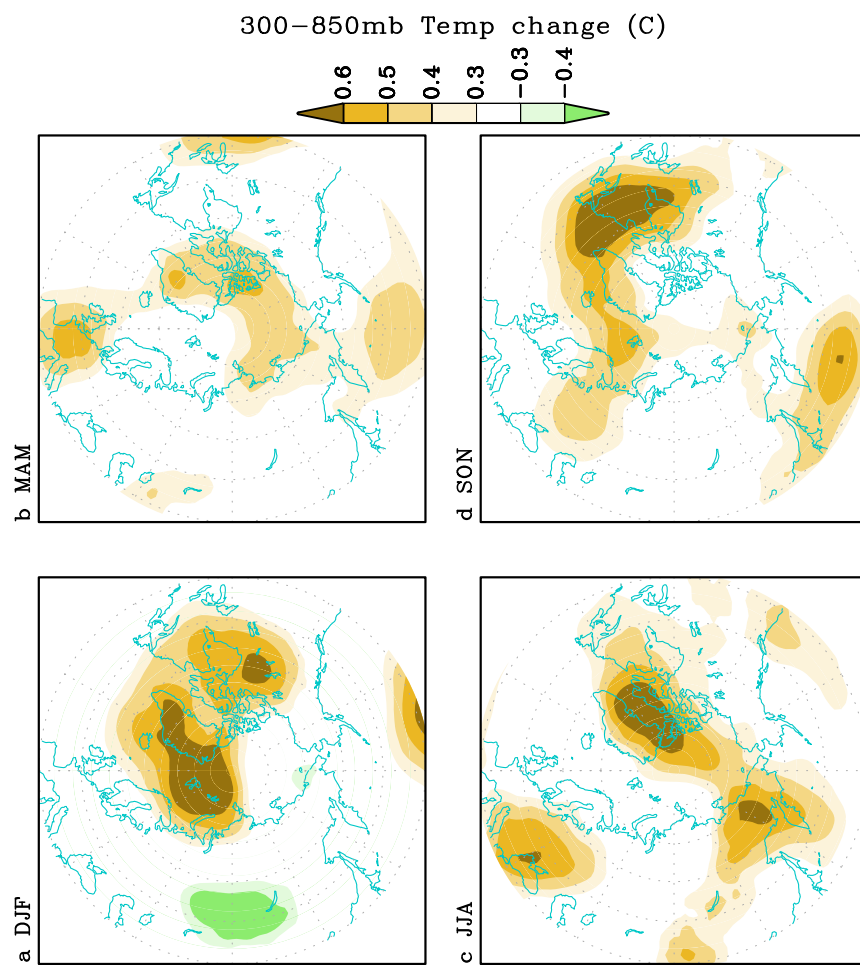
Extended Data Figure 1 | Observed trend pattern of annual mean surface temperature for 1979–2012. Linear trend (per decade) of annual mean surface and near-surface temperature for **a**, ERA-interim, **b**, MERRA reanalysis,

c, University of Delaware and **d**, GISTEMP for the period of 1979–2012. The grey regions in **c** and **d** indicate no data.



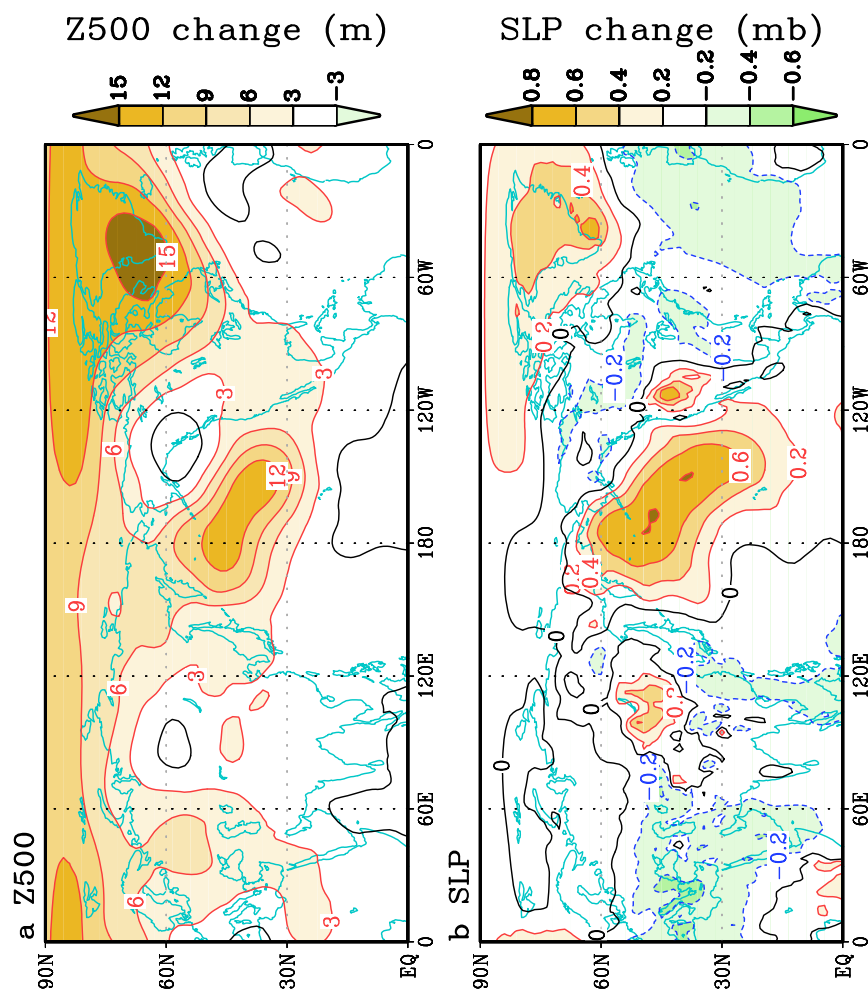
Extended Data Figure 2 | Observed trend pattern of seasonal mean surface temperature for 1979–2012. Linear trend (per decade) of seasonal mean ERA-interim surface temperature for **a**, December to February (DJF), **b**, March

to May (MAM), **c**, June to August (JJA) and **d**, September to November (SON) for the period 1979–2012.



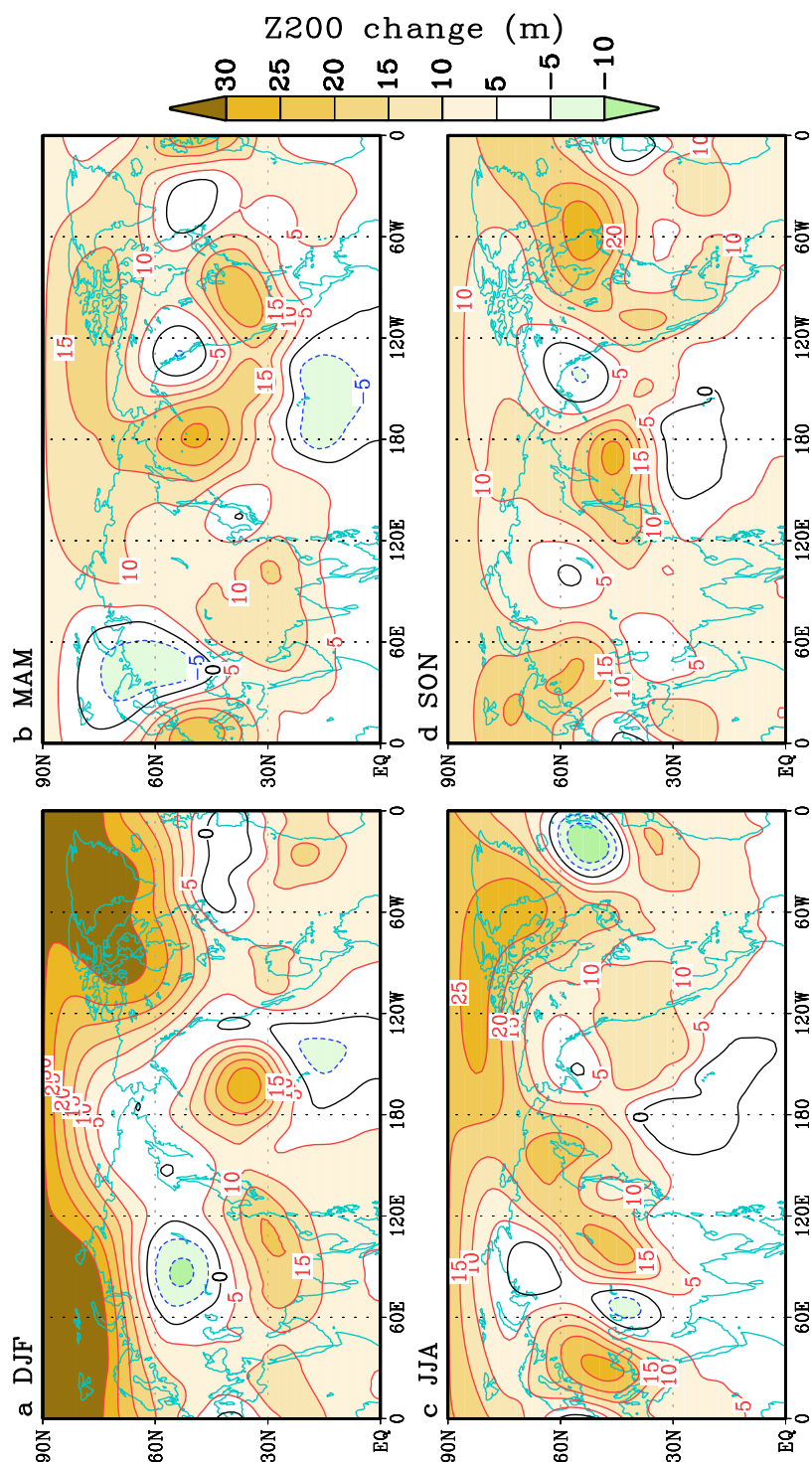
Extended Data Figure 3 | Observed trend pattern of seasonal mean 300–850 hPa temperature for 1979–2012. Linear trend (per decade) of

seasonal mean ERA-interim 300–850 hPa temperature for **a**, DJF, **b**, MAM, **c**, JJA and **d**, SON for the period 1979–2012.



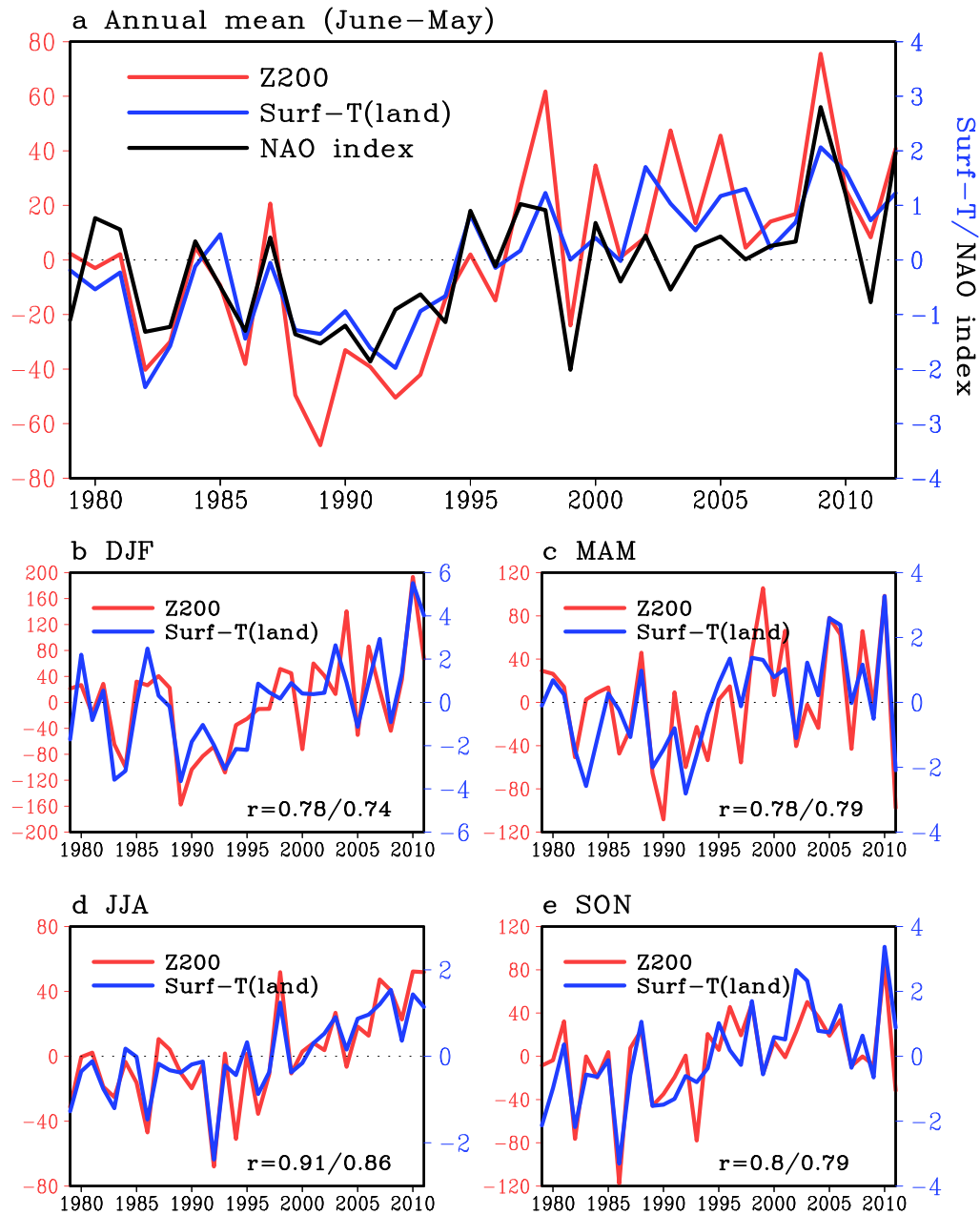
Extended Data Figure 4 | Observed trend pattern of annual mean field for 1979–2012. Linear trend (per decade) of annual mean **a**, 500-hPa geopotential

height (Z500), and **b**, sea level pressure (SLP) of ERA-interim for the period 1979–2012. Solid (dashed) contours indicate positive (negative) trends.



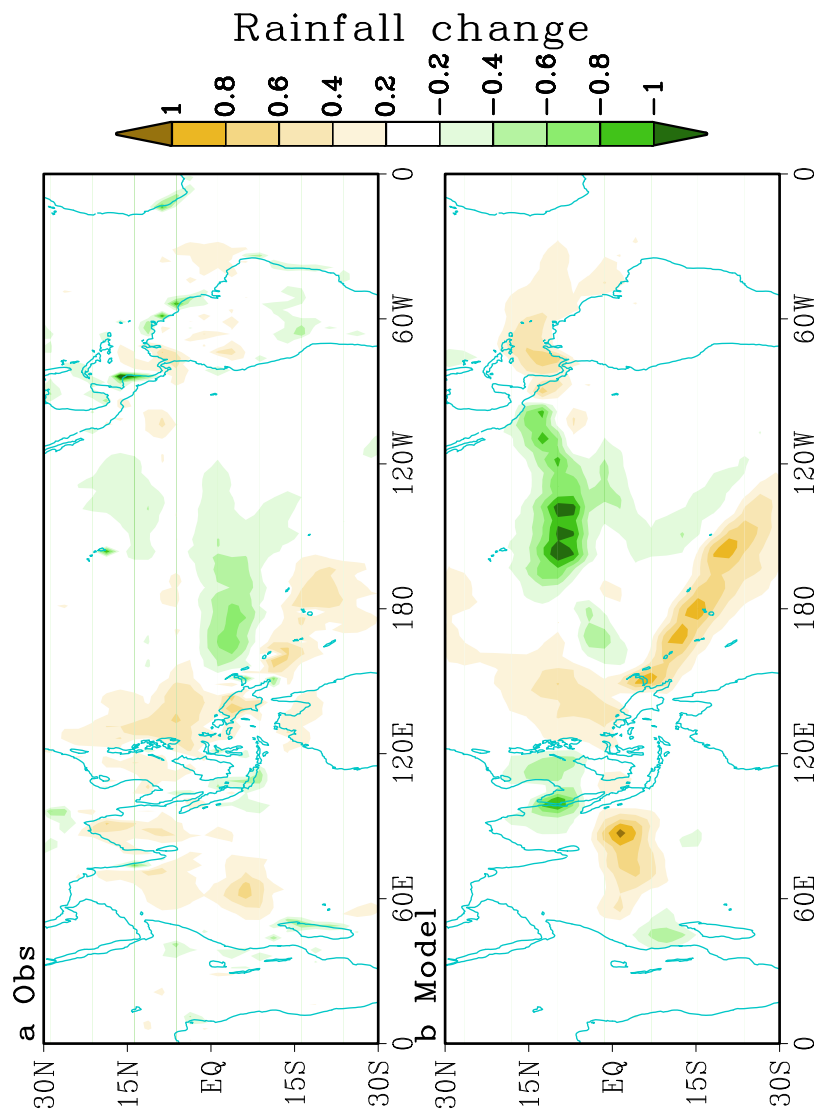
Extended Data Figure 5 | Observed trend pattern of seasonal mean 200-hPa geopotential height for 1979–2012. Linear trend (per decade) of seasonal mean ERA-interim 200-hPa geopotential height for a, DJF, b, MAM, c, JJA and

d, SON for the period 1979–2012. Solid (dashed) contours indicate positive (negative) trends.



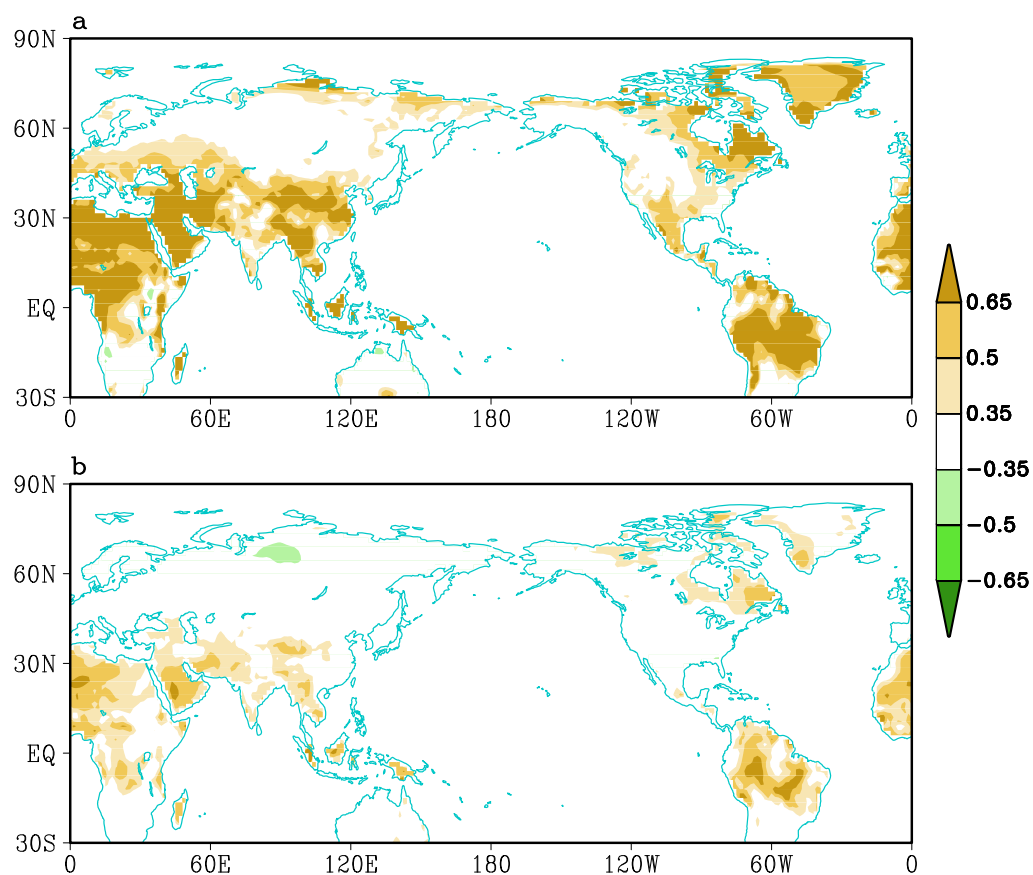
Extended Data Figure 6 | Observed surface temperature and Z200 change in northeastern Canada and Greenland. **a**, Annual mean surface temperature, Surf-T(land), and Z200 averaged over northeastern Canada and Greenland (55°N – 85°N , 280°E – 340°E , denoted by a box in Fig. 1c), and NOAA NAO index¹⁴ (sign is reversed for simplicity of comparison) for the period 1979–2012. The units on the left ordinate are geopotential metres; the units on

the right ordinate are $^{\circ}\text{C}$ (surface temperature) and unitless (NAO index). **b–e**, Surface temperature and Z200 averaged over northeastern Canada and Greenland (55°N – 85°N , 280°E – 340°E , denoted by a box in Fig. 1c) for each season for the period 1979–2012. The correlation coefficient (r) of two curves is denoted in the bottom right corner. The first (second) number denotes the correlation between the raw (detrended) time series.



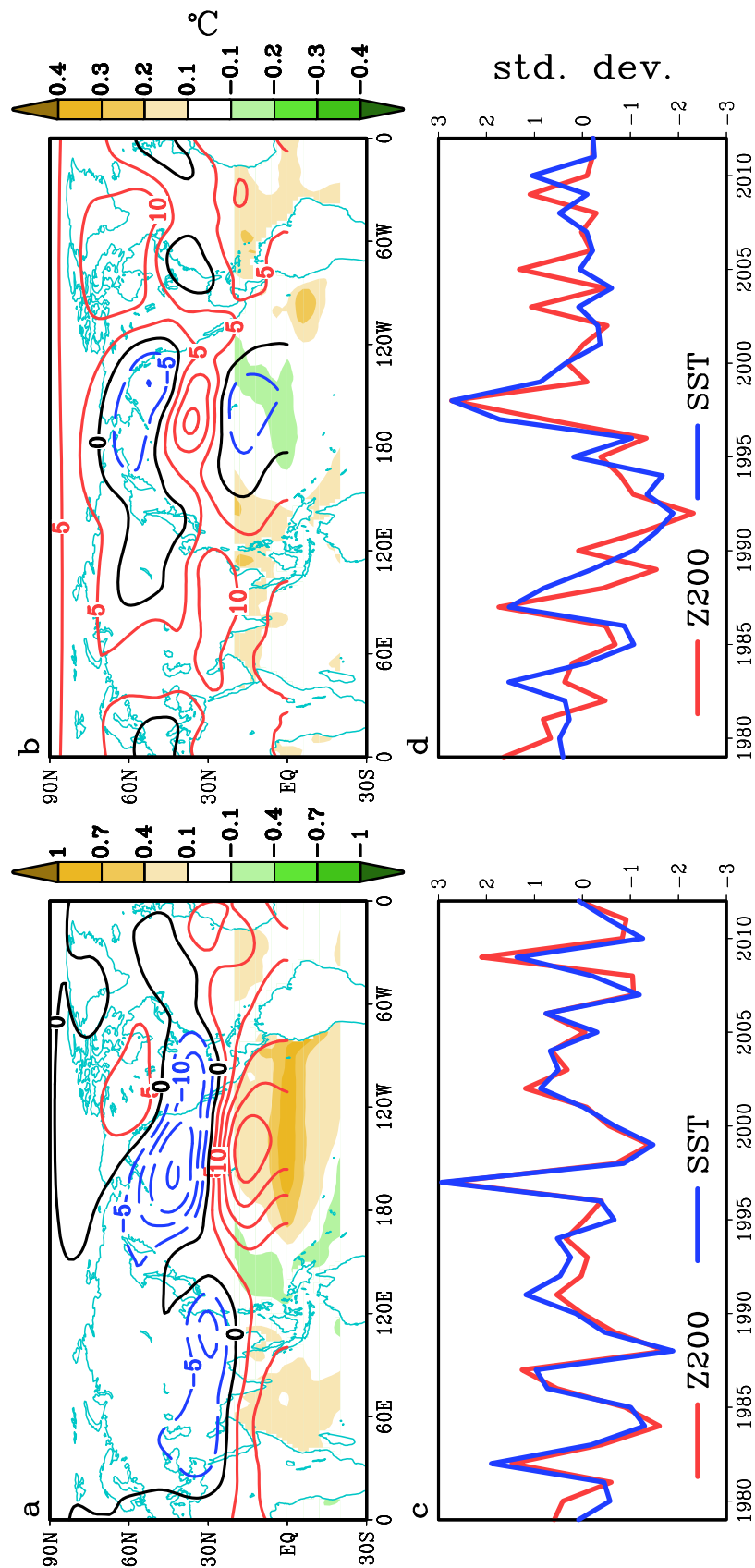
Extended Data Figure 7 | Observed and simulated annual mean tropical rainfall trend patterns for 1979–2012. Linear trend (mm per day per decade) of annual mean **a**, observed GPCP³¹ rainfall, **b**, simulated rainfall from 34-year

simulation of ECHAM run forced by observed SST (1979–2012) in the tropics (30° N to 30° S). In the extratropics and the polar region, the atmosphere is coupled to a slab ocean model with a thermodynamic sea-ice component.



Extended Data Figure 8 | MCA2 SST related annual mean land surface temperature. **a**, Correlation (colour scale shows r) between the time series of the MCA2 pattern of annual mean SST (blue curve in Fig. 2d) and annual mean

land surface temperature in ERA-interim during 1979–2012. **b**, As **a** but using the detrended surface air temperature data and the detrended time series of mode 2 SST.



Extended Data Figure 9 | Coupled patterns between detrended annual mean tropical SST and Northern Hemisphere circulation for 1979–2012.

Figure shows MCA results for detrended annual mean 1979–2012 Northern Hemisphere (0–88.5°N) 200-hPa geopotential heights (Z200) and tropical (20°S to 20°N) sea surface temperature (SST). Shown in **a** are the patterns of Z200 (contour interval 10 metres) and tropical SST (shading) that accompany

the first mode in these two data sets. Panel **c** displays the time series of the Z200 (red) and SST (blue) patterns shown in **a**. Panels **b** and **d** are the same as in **a** and **c** but for the second mode. Amplitudes in **a** and **b** are scaled by one standard deviation of the corresponding time series in **c** and **d**; the latter have a standard deviation of one.

Extended Data Table 1 | 40 climate models in the CMIP5 historical experiment

CMIP5 model designation	nx	ny
1. bcc-csm1-1	128	64
2. bcc-csm1-1-m	320	160
3. CCSM4	288	192
4. CESM1-BGC	288	192
5. CESM1-CAM5	288	192
6. CESM1-FASTCHEM	288	192
7. CESM1-WACCM	144	96
8. CMCC-CESM	96	48
9. CMCC-CM	480	240
10. CMCC-CMS	192	96
11. CNRM-CM5	256	128
12. CSIRO-Mk3-6-0	192	96
13. CanESM2	128	64
14. FGOALS-g2	128	60
15. FGOALS-s2	128	108
16. FIO-ESM	128	64
17. GFDL-CM2p1	144	90
18. GFDL-CM3	144	90
19. GFDL-ESM2G	144	90
20. GFDL-ESM2M	144	90
21. GISS-E2-H	144	89
22. GISS-E2-H-CC	144	89
23. GISS-E2-R	144	89
24. GISS-E2-R-CC	144	89
25. HadCM3	96	73
26. HadGEM2-AO	192	144
27. Inmcm4	180	120
28. IPSL-CM5A-LR	96	96
29. IPSL-CM5A-MR	144	143
30. IPSL-CM5B-LR	96	96
31. MIROC-ESM	128	64
32. MIROC-ESM-CHEM	128	64
33. MIROC4h	640	320
34. MIROC5	256	128
35. MPI-ESM-LR	192	96
36. MPI-ESM-MR	192	96
37. MPI-ESM-P	192	96
38. MRI-CGCM3	320	160
39. NorESM1-M	144	96
40. NorESM1-ME	144	96

List of 40 CMIP5 CGCMs used in Fig. 4 to examine the forced response of the climate system to anthropogenic and natural external forcing, along with the number of atmospheric horizontal grids.

Predicting biodiversity change and averting collapse in agricultural landscapes

Chase D. Mendenhall^{1,2}, Daniel S. Karp^{1,2,3,4}, Christoph F. J. Meyer^{5,6}, Elizabeth A. Hadly² & Gretchen C. Daily^{1,2,7,8,9}

The equilibrium theory of island biogeography¹ is the basis for estimating extinction rates² and a pillar of conservation science^{3,4}. The default strategy for conserving biodiversity is the designation of nature reserves, treated as islands in an inhospitable sea of human activity⁵. Despite the profound influence of islands on conservation theory and practice^{3,4}, their mainland analogues, forest fragments in human-dominated landscapes, consistently defy expected biodiversity patterns based on island biogeography theory^{6–13}. Countryside biogeography is an alternative framework, which recognizes that the fate of the world's wildlife will be decided largely by the hospitality of agricultural or countryside ecosystems^{12,14–17}. Here we directly test these biogeographic theories by comparing a Neotropical countryside ecosystem with a nearby island ecosystem, and show that each supports similar bat biodiversity in fundamentally different ways. The island ecosystem conforms to island biogeographic predictions of bat species loss, in which the water matrix is not habitat. In contrast, the countryside ecosystem has high species richness and evenness across forest reserves and smaller forest fragments. Relative to forest reserves and fragments, deforested countryside habitat supports a less species-rich, yet equally even, bat assemblage. Moreover, the bat assemblage associated with deforested habitat is compositionally novel because of predictable changes in abundances by many species using human-made habitat. Finally, we perform a global meta-analysis of bat biogeographic studies, spanning more than 700 species. It generalizes our findings, showing that separate biogeographic theories for countryside and island ecosystems are necessary. A theory of countryside biogeography is essential to conservation strategy in the agricultural ecosystems that comprise roughly half of the global land surface and are likely to increase even further¹⁴.

The mark of humanity on the biosphere is now so widespread and profound that many consider Earth to have entered a new geological epoch, the Anthropocene. This new epoch has brought an explosion of countryside ecosystems—arenas where villages, farmland and minimally altered habitats weave together to form intricate, human-dominated ecosystems—where the fate of much of the planet's wildlife is being determined^{14,15}. Driven by the unrelenting pace of biodiversity loss and the erosion of benefits that nature provides to society¹⁸, conservation science has entered a new phase of prolific innovation¹⁹. Despite the broadly defined and rapidly adapting framework of conservation science²⁰, many of the biological theories at its centre persist, relatively unchanged. For example, theories from simple island ecosystems are still used in ways that incorrectly estimate rates of species extinction²¹ and distort projections of ecological risk in human-dominated ecosystems^{2,22}, further exhausting an environmental, apocalyptic narrative²³. Confusion and controversy surrounding the biodiversity crisis are substantially related to the overextension of the theory of island biogeography to human-dominated ecosystems, exacerbating a rift between conservation theory and real-world practices^{20,24}.

More than 30 years after the establishment of experiments⁶ designed to test the efficacy of island biogeographic theory in countryside ecosystems, we now know that wildlife often defies patterns predicted by island biogeography theory—emphasizing the critical role of human-made habitats in ecosystems^{11,14–16,21,24}. Despite its known shortcomings^{10,11}, this theory persists as the basis for estimating extinction rates and making policy recommendations^{2,21,22}. In response, a new, holistic view of biogeography is emerging for countryside ecosystems—one that incorporates human-made habitats and their associated biodiversity^{14–17,24,25}.

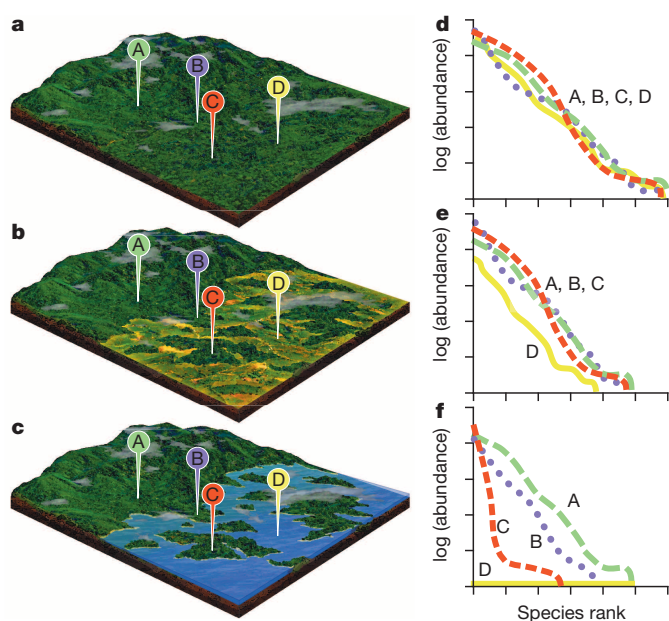
Here we predict that human-made habitats typical of many countryside ecosystems will provide resources for many species to increase their abundances and mitigate the extirpation of populations and species. We propose three hypotheses to contrast countryside and island biogeographic theories empirically: first, that forest fragments in agricultural landscapes, in comparison with true islands, support more species and have lower rates of local extinction; second, that forest fragments surrounded by deforested habitat host more evenly distributed abundances of species than do true islands surrounded by water because human-made habitats support individuals to differing degrees, through part or all of their life cycles; and third, that the expansion by many forest species into human-made habitats in the countryside ecosystem is generating novel assemblages of biodiversity, whose members and their abundances reflect species-specific responses to habitat change (Fig. 1).

We test these hypotheses by using data on Neotropical bat biodiversity, a group acutely sensitive to tropical deforestation⁹. First, we contrast bat biodiversity within a mosaic of forest fragments embedded in farmland with that on true islands within a large, human-made lake. Second, we perform a meta-analysis of 29 studies of bat species richness across countryside and island ecosystems worldwide, to assess the generality of our findings.

Specifically, we contrast bat biodiversity in a Costa Rican countryside dominated by coffee plantations and pastures with a Panamanian island ecosystem dominated by water (Fig. 2). The two ecosystems are ideal for testing biogeographic theories because of their approximately similar ages, their close geographic proximity, their evolutionarily homologous bat diversity and their similarly sized historical biogeographic species pools.

Bats were captured over a 2–4-year period in sites located in forest reserves, forest fragments and coffee plantations in the countryside ecosystem and in mainland forests, mainland forest edges and true islands in the island ecosystem²⁶ (Fig. 2; see Methods). We captured 9,283 individual bats (4,282 and 5,001 individuals in the countryside and island ecosystem, respectively) and 62 bat species (43 and 33 species in the countryside and island ecosystem, respectively; 14 species were shared between the ecosystems). We found striking differences between ecosystems in three dimensions of biodiversity corresponding to our three hypotheses: species richness, species evenness and abundance-based assemblage composition (Fig. 3 and Extended Data Tables 1–3; see Methods).

¹Center for Conservation Biology, Stanford University, Stanford, California 94305, USA. ²Department of Biology, Stanford University, Stanford, California 94305, USA. ³Department of Environmental Science, Policy & Management, University of California, Berkeley, California 94720, USA. ⁴The Nature Conservancy, Berkeley, California 94705, USA. ⁵Institute of Experimental Ecology, University of Ulm, 89069 Ulm, Germany. ⁶Centre for Environmental Biology, University of Lisbon, 1749-016 Lisbon, Portugal. ⁷Woods Institute for the Environment, Stanford University, Stanford, California 94305, USA. ⁸Global Economic Dynamics and the Biosphere, Royal Swedish Academy of Sciences, Stockholm SE-104 05, Sweden. ⁹Stockholm Resilience Centre, University of Stockholm, Stockholm SE-106 91, Sweden.



To test our first hypothesis, we compared estimated species richness in forest fragments with that in true islands (Fig. 3b). In agreement with the species–area relationships that include human-made habitats^{12,13}, the best-fit statistical models that explained bat species richness in forest reserves and fragments in the countryside ecosystem suggested either no change in species richness between habitats (mean species estimated

Figure 1 | Hypothetical biodiversity changes in countryside and island ecosystems. **a–c**, Conceptual diagrams showing hypothesized patterns of biodiversity in a minimally altered ecosystem (**a**) and corresponding countryside (**b**) and island (**c**) ecosystems derived from hypothetical land conversion either to agriculture or a lake, respectively. **d–f**, Proposed patterns of biodiversity for minimally altered (**d**), countryside (**e**) and island (**f**) ecosystems, depicted by rank–abundance plots. In each rank–abundance plot, the most abundant species at each site is given a species rank of 1, the second most abundant is 2, and so on. Pin labels mark sites and habitats with letters in each ecosystem and are referenced in the rank–abundance plot. In the countryside and island ecosystems, letters represent sites located in the mainland or reserve interior (A), mainland or reserve edge (B), forest fragment or true island (C) and deforested habitat or water matrix (D). We speculate that species loss generally occurs after landscape alteration, but there is wide variation between countryside and island ecosystems in how biodiversity changes in terms of species loss, changes in abundances and the formation of novel assemblages of biodiversity among habitats, because of the resources afforded in human-made habitats in countryside ecosystems.

per site = 25 (s.d. = 4); Extended Data Table 4) or little change (competing model with a difference in the corrected Akaike information criterion (ΔAIC_c) = 0.220 yielded a mean species estimated per reserve site of 28 (s.d. = 5) and a mean species estimated per forest fragment site of 23 (s.d. = 3)). Fragment size and isolation distance were poor predictors of species richness among countryside forest reserves and fragments (Extended Data Table 3).

Aligning with our first hypothesis and predictions of island biogeographic theory, the estimated number of bat species in the island ecosystem declined with isolation distance but increased with island size (Extended Data Table 4). The area of remaining tropical forest within 1 km, a more ecologically relevant measure of habitat amount¹¹, was the

Sampling sites and habitats

- ▲ Reserve
- Forest fragment
- ◆ Coffee plantation
- ▲ Mainland
- Mainland edge
- Island
- Forest habitat
- Deforested habitat
- Water

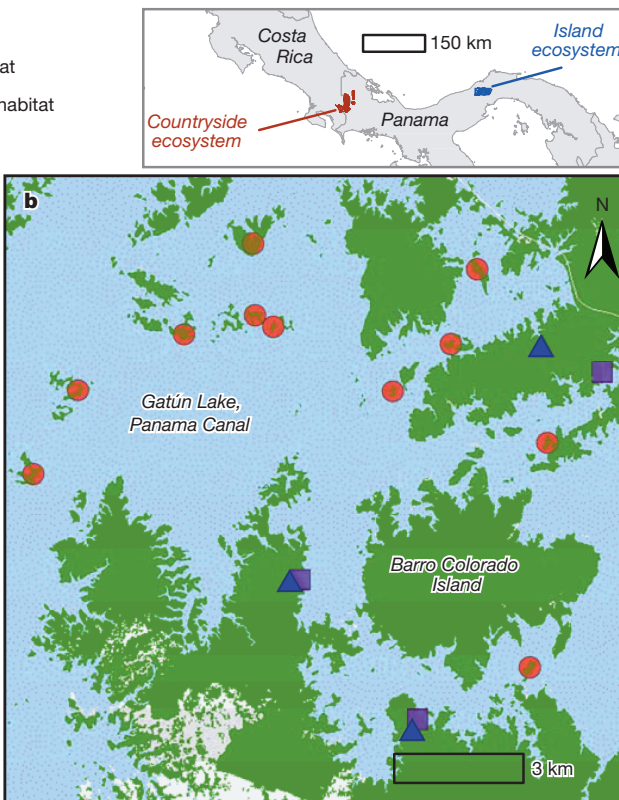
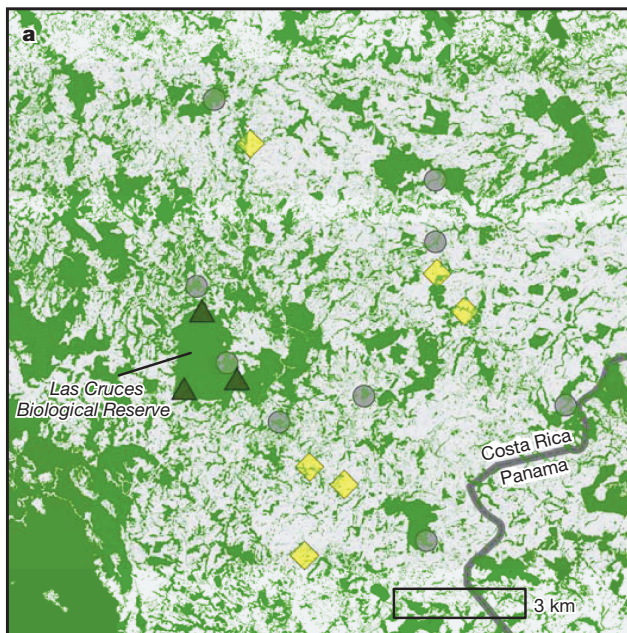
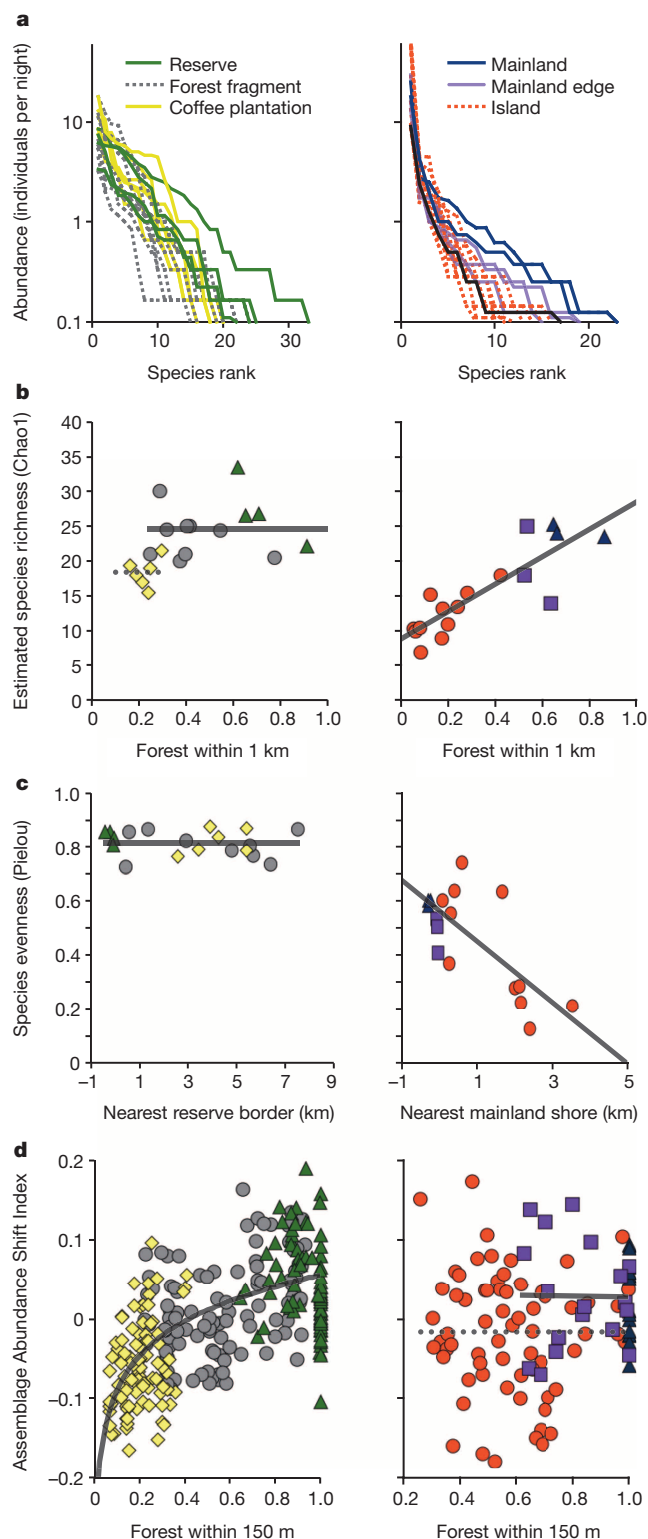


Figure 2 | Bat sampling locations. A total of 36 locations were sampled for bat biodiversity: 19 across the countryside ecosystem located in southern Costa Rica (**a**) and 17 in the human-made island ecosystem of Gatún Lake, constituting a major part of the Panama Canal (**b**). The two ecosystems are separated by about 350 km and occupy the neighbouring tropical life zones of premontane wet and lowland moist forests. Both ecosystems were extensively sampled for bats by using ground-level mist nets over multiple years. In the countryside ecosystem, sampling was conducted in two different forest

reserves, a variety of forest fragments, and coffee plantations. In the island ecosystem, sampling was conducted in the interior mainland, mainland edges and on a variety of islands. The island ecosystem was sampled from 2003 to 2005 and the countryside ecosystem from 2009 to 2012 with similar sampling techniques and effort (see Methods). In the countryside ecosystem figure the sampling location inside La Amistad International Park is not shown but is located about 25 km northeast of the Las Cruces Biological Reserve of the Organization for Tropical Studies.



strongest predictor of species richness on islands ($R^2 = 0.746$, $P < 0.001$, $n = 17$; Fig. 3a and Extended Data Table 4). In support of our first hypothesis, the rate of species loss as forest area declined was substantially and significantly higher in the island ecosystem than in the countryside ecosystem (Extended Data Table 2).

To test our second hypothesis, we analysed Pielou's evenness index (Fig. 3c). Forest reserves and fragments in the countryside ecosystem did not vary in evenness—probably because bat populations are supported beyond forest edges in the deforested habitat. In the island ecosystem,

Figure 3 | Countryside and island bat biodiversity patterns. Countryside and island ecosystems support bat biodiversity in fundamentally different ways. A total of 62 bat species were recorded: 43 were captured in the countryside ecosystem (left column) and 33 in the island ecosystem (right column); of these, 14 bat species were shared. Shown are the countryside and island ecosystem rank–abundance distributions for each site (a) and patterns of bat species richness (b), species evenness (c) and the Assemblage Abundance Shift Index (d), which accounts for changes in species richness. Symbol shapes and colours correspond to those shown in Fig. 2. Lines depict best-fit relationships after model selection and optimization. Dotted lines correspond to coffee plantations in b and islands in d, where broad habitat type was the best predictor. Analyses in b and c were conducted at a site level (36 sites). Species richness was estimated to account for rare species using the Chao1 method and species evenness is standardized between 0 and 1 using the Pielou method. Analysis in d was conducted at a net level (383 nets), because assemblage differences were observed within sites relative to fine-scale changes in forest cover at 150 m. See Methods for further details, and Extended Data Tables 1–4 for model comparisons and regression coefficients.

however, species abundances were increasingly uneven on more isolated islands (Fig. 3c and Extended Data Table 5). Uneven abundance distributions of bats in the island ecosystem are probably explained by island biogeography's faunal density compensation theory²⁷.

To test our third hypothesis, we examined bat biodiversity patterns in a variety of coffee plantations in the countryside ecosystem (Fig. 3). In coffee plantations we captured 1,508 individuals representing 26 of the 43 bat species (mean species estimated per coffee plantation site = 18 (s.d. = 2); Fig. 3b and Extended Data Table 4), nine of which were observed roosting and/or breeding on farmland but not in pastures²⁸ (Extended Data Table 6). Moreover, bat species in coffee plantations were evenly distributed in their abundances (Fig. 3c). Species varied considerably in their abundance responses to deforestation, but 14 of the 30 bat species with at least seven captures in the countryside ecosystem were more abundant outside the forest reserves. Further, no species showed complete dependence on a single habitat type (Extended Data Fig. 1).

Finally, to evaluate whether a novel bat assemblage is forming in coffee plantations because of predictable, species-specific changes in abundances, we developed and compared an index that measured assemblage-level changes in bat abundances across habitats in both ecosystems. The Assemblage Abundance Shift Index is based on ordination analyses of bat abundances and how they shift collectively, relative to bat assemblages in forest reserves in the countryside ecosystem and to mainland sites in the island ecosystem. The Assemblage Abundance Shift Index accounts for changes in species richness and detection bias (see Methods and Extended Data Fig. 2). Patterns of the Assemblage Abundance Shift Index revealed that the abundances of many bat species in the countryside ecosystem underwent predictable transitions in tandem with declining forest cover, particularly at small spatial scales¹¹ (150 m) and especially in coffee plantations. In the island ecosystem, bat abundances differed significantly between assemblages on the mainland interior and edge sites in comparison with islands, probably because of density compensation²⁷ (Fig. 3d).

Our meta-analysis found strong evidence that bat biodiversity patterns follow predictions of island biogeographic theory on true islands—both natural and human-made—but countryside ecosystems are more complex. A review of 206 peer-reviewed papers on bat biodiversity in island and countryside ecosystems yielded several key findings from the 29 studies that met meta-analysis search criteria (Fig. 4a; see Methods). First, we confirmed that the equilibrium theory of island biogeography accurately forecasted the change in bat species richness with island size on true islands (Fig. 4b), and bats on islands followed typical species–area relationships (Fig. 4c). Second, we found that declines in bat species richness expected from island biogeography were almost never realized in countryside forest fragments (Fig. 4b). Third, we found a variety of bat species richness responses in open habitats of countryside ecosystems, including relatively high richness in some agricultural systems^{15,24} (Fig. 4d). Finally, we found that 10 of the 12 studies that compared bat assemblages between minimally altered forest, forest

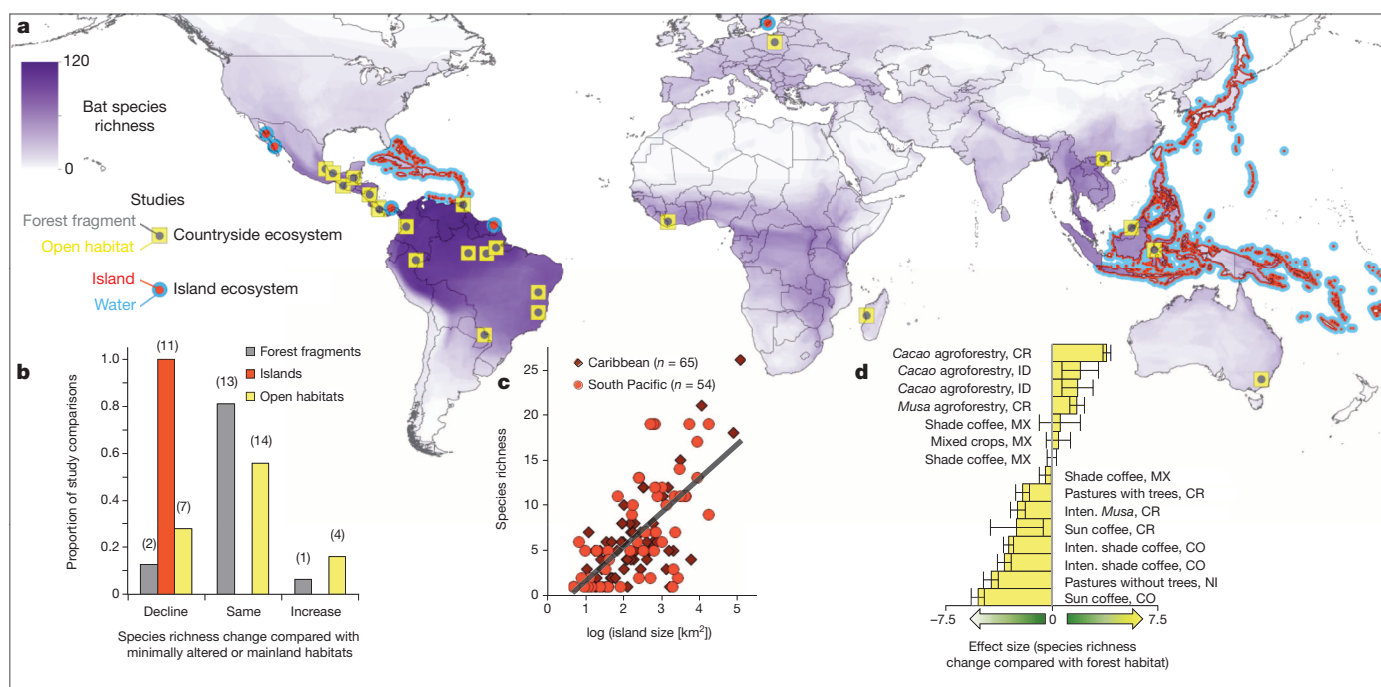


Figure 4 | Bat species richness responses in countryside and island ecosystems worldwide. Globally, countryside ecosystems and island ecosystems support biodiversity in fundamentally different ways, necessitating separate biogeographic theories for prediction and management. **a**, Map identifying the locations of 52 bat species richness comparisons in different countryside and island habitats gathered from 29 studies. In all, the studies surveyed more than 700 bat species, representing more than 60% of all bat and 10% of all mammal species globally. **b–d**, Island bat species richness always declined relative to the mainland or to larger islands in island ecosystems (**b**). For example, bat species richness on islands in the Caribbean and South Pacific yielded a typical species–area relationship (**c**; $S = cA^z$, $c = 0.043$, $z = 0.269$, where S is the number of species, A is the habitat area, c is the

fragments and/or human-made habitats reported a significant assemblage shift in bat biodiversity (see Supplementary Data).

In view of humanity's incomplete knowledge of nature, the designation of protected areas is, and will continue to be, an essential strategy for conserving biodiversity⁹. The widespread application of island biogeographic theory to countryside—less so in Europe, but especially in the Tropics—is, however, distorting our understanding and conservation strategies in agriculture, the enterprise on which the future of biodiversity hinges most critically. The hospitality of the world's agricultural lands is threatened by an increasing use of chemical inputs and practices that sterilize, structurally level, and standardize plots^{14,15,17,24}—homogenizing and decimating biodiversity²⁹. Here we provide evidence that a countryside biogeographic framework that is inclusive of human-made habitats and the opportunities it can afford to many species—given appropriate management of those habitats^{15,24}—better represents how, at least, bat biodiversity is responding in the Anthropocene. Approaches for estimating biodiversity that explicitly account for human-made habitats^{4,11–13,21,30} are critical for predicting biodiversity change and avoiding its collapse in the agricultural landscapes that will increasingly dominate the Earth.

METHODS SUMMARY

Bats were sampled by using constant-effort mist netting. In the Costa Rican countryside ecosystem, mist-netting protocols consisted of 20 ground-level mist nets (12 m × 2.5 m, 32-mm mesh) in a 3–5-ha plot haphazardly placed at each site, except the 'supersite' located in La Amistad International Park, which consisted of 60 haphazardly placed mist nets within a 10-ha plot. Constant-effort mist netting for bats in the countryside ecosystem took place between 24 January and 28 March for four years (2009–2012). Sampling began at sunset and lasted 4 h each night. Countryside ecosystem sites were sampled on three nights each year from 2009 to 2011

and on only one night in 2012, except the 'supersite' located in La Amistad International Park, which was sampled for three nights in 2009 and three nights in 2010. In the Panamanian island ecosystem, mist-netting protocols consisted of six ground-level mist nets (6 m × 2.5 m, 16-mm mesh) placed around the perimeter of a 0.5-ha plot at each site. Constant-effort mist netting for bats in the island ecosystem took place between October 2003 and October 2005. Each site in the island ecosystem was sampled for seven or eight complete nights (that is, from sunset to sunrise) over a 2-year period. All captured bats in both ecosystems were identified and standard measurements were collected. Adult bats from larger species (more than 5–10 g) were marked individually.

constant associated with the area measurement, and z is the slope of the species–area relationship in log–log space). In contrast, bat species richness in countryside forest fragments generally had the same number of species as minimally altered forest habitat (**b**). Bat species richness in open habitats compared to forest habitat varied (**b**). Studies with enough information to calculate effect sizes demonstrated how some agricultural practices, such as agroforestry, support more bat biodiversity than more intensive agricultural practices (**d**; effect size direction indicates species richness decline or increase compared to forest habitat, error bars represent 95% confidence intervals). Inten., intensive. Country code abbreviations: CR, Costa Rica; CO, Colombia; ID, Indonesia; MX, Mexico.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 December 2013; accepted 10 February 2014.

Published online 16 April; corrected online 7 May 2014 (see full-text HTML version for details).

- MacArthur, R. & Wilson, E. *The Theory of Island Biogeography* 203 (Princeton Univ. Press, 1967).
- Millennium Ecosystem Assessment. *Ecosystems and Human Well-being* (Island Press and World Resources Institute, 2005).
- Simberloff, D. & Abele, L. Island biogeography theory and conservation practice. *Science* **191**, 285–286 (1976).
- Hanski, I. & Gilpin, M. Metapopulation dynamics. *Biol. J. Linn. Soc.* **42**, 3–16 (1991).
- Wilson, E. & Willis, E. in *Ecology and Evolution of Communities* (eds Cody, M. L. & Diamond, J. M.) 522–534 (Belknap, 1975).
- Laurance, W. F. *et al.* Ecosystem decay of Amazonian forest fragments. *Conserv. Biol.* **16**, 605–618 (2002).
- Debinski, D. & Holt, R. A survey and overview of habitat fragmentation experiments. *Conserv. Biol.* **14**, 342–355 (2000).
- Laurance, W. Theory meets reality: How habitat fragmentation research has transcended island biogeographic theory. *Biol. Conserv.* **141**, 1731–1744 (2008).
- Laurance, W. *et al.* Averting biodiversity collapse in tropical forest protected areas. *Nature* **489**, 290–294 (2012).

10. Lomolino, M. A call for a new paradigm of island biogeography. *Glob. Ecol. Biogeogr.* **9**, 1–6 (2000).
11. Fahrig, L. Rethinking patch size and isolation effects. *J. Biogeogr.* **40**, 1649–1663 (2013).
12. Pereira, H. & Daily, G. Modeling biodiversity dynamics in countryside landscapes. *Ecology* **87**, 1877–1885 (2006).
13. Guilherme, J. & Pereira, H. Adaptation of bird communities to farmland abandonment in a mountain landscape. *PLoS ONE* **8**, e73619 (2013).
14. Mendenhall, C., Kappel, C. & Ehrlich, P. in *Encyclopedia of Biodiversity* (ed. Levin, S.) 347–360 (Elsevier, 2013).
15. Perfecto, I., Vandermeer, J. & Wright, A. *Nature's Matrix: Linking Agriculture, Conservation and Food Sovereignty* 242 (Earthscan, 2009).
16. Daily, G. Ecological forecasts. *Nature* **411**, 245 (2001).
17. Daily, G., Ceballos, G., Pacheco, J., Suzán, G. & Sánchez-Azofeifa, A. Countryside biogeography of Neotropical mammals. *Conserv. Biol.* **17**, 1814–1826 (2003).
18. Cardinale, B. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
19. Kueffer, C. & Kaiser-Bunbury, C. Reconciling conflicting perspectives for biodiversity conservation in the Anthropocene. *Front. Ecol. Environ.* **10**, 1890/120201 (2013).
20. Kareiva, P. & Marvier, M. What is conservation science? *Bioscience* **62**, 962–969 (2012).
21. He, F. & Hubbell, S. Species–area relationships always overestimate extinction rates from habitat loss. *Nature* **473**, 368–371 (2011).
22. Gibson, L. *et al.* Near-complete extinction of native small mammal fauna 25 years after forest fragmentation. *Science* **341**, 1508–1510 (2013).
23. Heise, U. *Sense of Place and Sense of Planet: The Environmental Imagination of the Global* 250 (Oxford Univ. Press, 2008).
24. Tscharntke, T. *et al.* Global food security, biodiversity conservation and the future of agricultural intensification. *Biol. Conserv.* **151**, 53–59 (2012).
25. Rosenzweig, M. *Win–Win Ecology: How the Earth's Species Can Survive in the Midst of Human Enterprise* 211 (Oxford Univ. Press, 2003).
26. Meyer, C. & Kalko, E. Assemblage-level responses of phyllostomid bats to tropical forest fragmentation. *J. Biogeogr.* **35**, 1711–1726 (2008).
27. MacArthur, R., Diamond, J. & Karr, J. Density compensation in island faunas. *Ecology* **53**, 330–342 (1972).
28. Reid, J., Holste, E. & Zahawi, R. Artificial bat roosts did not accelerate forest regeneration in abandoned pastures in southern Costa Rica. *Biol. Conserv.* **167**, 9–16 (2013).
29. Karp, D. *et al.* Intensive agriculture erodes β -diversity at large scales. *Ecol. Lett.* **15**, 963–970 (2012).
30. Rybicki, J. & Hanski, I. Species–area relationships and extinctions caused by habitat loss and fragmentation. *Ecol. Lett.* **16**, 27–38 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank P. Ehrlich, E. Kalko, F. Oviedo Brenes, R. Zahawi, L. Frishkoff, K. Holl, H. Kim Frank, M. Knope, J. L. Reid, A. Wrona, H. York and dozens of field assistants and Costa Rican landowners, and the communities and staffs of the Organization for Tropical Studies, Las Cruces Biological Station, the Smithsonian Tropical Research Institute and the Center for Conservation Biology at Stanford University. Research was funded by the Winslow Foundation, the Moore Family Foundation, the German Academic Exchange Service, the German Science Foundation, Peter and Helen Bing, Ralph and Louise Haberfeld, and a Restoration Workshop Research Grant through the Las Cruces Biological Station. C.D.M. and D.S.K. were supported by National Science Foundation Graduate Research Fellowships.

Author Contributions C.D.M. and G.C.D. conceived the study. C.D.M. collected data from Costa Rica, performed analyses, and wrote the manuscript. C.F.J.M. collected data from Panama. D.S.K. assisted with key elements of analysis. All authors contributed ideas to the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to C.D.M. (cdm@stanford.edu).

METHODS

Costa Rican countryside ecosystem bat sampling. The countryside investigation was centred on the Las Cruces Biological Station and Wilson Botanical Garden located in the Coto Brus Valley in southwestern Costa Rica. The Las Cruces Biological Reserve (8° 47' N, 82° 57' W, 1,100 m above sea level), which serves as a regional biological baseline, protects about 280 ha of primary and mature secondary forest (Fig. 2). The climate is characterized by a long rainy season (9–10 months) followed by a brief dry season, with an average annual rainfall of 3.4 m (ref. 31). Forests in the area are classified as premontane tropical wet forest³². More than 60% of the Coto Brus Valley, once forested, has been converted into cropland and pasture since the 1950s (refs 33, 34). Despite relatively recent deforestation, sediment core analyses reveal a robust, 3,000-year history of agriculture, forest clearing and fire in the region up to the sixteenth century^{35,36}.

Bats were sampled in seven general habitat types at 19 locations within 25 km of Las Cruces Biological Reserve (Fig. 2). All except one habitat type were represented by three sites located in (1) the Las Cruces Biological Reserve (three sites, located 1.3, 1.8 and 1.8 km apart from their nearest neighbouring site in the reserve), (2) small forest fragments (1–20 ha, three sites), (3) 10–20-year-old secondary forest located in large forest fragments (50–75 ha, three sites), (4) thin strips of riparian remnant forests (10–100 m wide, three sites), (5) coffee plantations with about 25% local tree cover on farms (three sites) and (6) coffee plantations with about 5% local tree cover (three sites). Moreover, one 'supersite', or a site equivalent in sampling effort to three sites (that is, 60 mist nets were deployed, instead of the normal 20 mist nets at the other 18 sites), operated in La Amistad International Park located about 25 km northeast of the Las Cruces Biological Reserve of the Organization for Tropical Studies. Bat sampling was conducted using constant-effort mist netting. Mist-netting protocols consisted of 20 ground-level mist nets (12 m × 2.5 m, 32-mm mesh) in a 3–5-ha plot haphazardly placed at each site, except the 'supersite' located in La Amistad International Park, which consisted of 60 haphazardly placed mist nets within a 10-ha plot. Constant-effort mist netting for bats took place between 24 January and 28 March for four years (2009–2012). Sampling began at sunset and lasted for 4 h each night. Sites were sampled on three nights each year from 2009 to 2011 and only one night in 2012, except the 'supersite' located in La Amistad International Park, which was sampled for three nights in 2009 and three nights in 2010.

Captured bats were identified and standard measurements were collected by the method of ref. 37. In 2009, bats were individually, but temporarily, marked by using wing punches³⁹. From 2010–2012 most bats (species weighing more than 5 g, excluding juveniles) were marked with collars fashioned from self-locking cable ties and individually numbered aluminium bands³⁷. All bats were released on site shortly after identification, marking and measurement. All animals were handled and treated ethically within the guidelines of the Institutional Animal Care and Use Committee.

Panamanian island ecosystem bat sampling. The island investigation was conducted on the human-made Gatún Lake (9° 11' N, 79° 53' W, 26 m above sea level) located in Central Panama. The 1914 damming of the Chagres River during construction of the Panama Canal created the lake and more than 200 islands ranging in size from less than 1 ha to the well-studied 1,560-ha Barro Colorado Island (BCI). Together with five adjacent mainland peninsulas, BCI forms the 5,400-ha Barro Colorado Nature Monument and is contiguous with 22,000-ha Soberanía National Park. The climate is strongly seasonal, with a long rainy season followed by a 4-month dry season. Average annual rainfall is 2.1 m (ref. 38). Dry-season winds have affected exposed islands, resulting in lower forest stature and tree species diversity⁴⁰. Forests in the area are classified as lowland tropical moist forest³².

Bats were sampled at six sites on the mainland peninsulas and at eleven sites on islands of various sizes and distances from the mainland. Bat sampling was conducted with constant-effort mist netting. Mist-netting protocols consisted of six ground-level mist nets (6 m × 2.5 m, 16-mm mesh) placed around the perimeter of a 0.5-ha plot at each site. Constant-effort mist netting for bats took place between October 2003 and October 2005. Each site was sampled for seven or eight complete nights (that is, sunset to sunrise) over a 2-year period. Bats were also captured with a single, large wall net. Bats captured in the wall net were excluded from this study because comparable sampling of the forest canopy was not conducted in the Costa Rican countryside ecosystem.

Captured bats were identified, and standard measurements and demographic data were collected as described in ref. 41. Most bats (species weighing more than 10 g, excluding juveniles) were marked with individually numbered ball-chain necklaces, or, in the case of larger species, with passive, subcutaneous transponders (EURO-ID). All bats were released on site shortly after marking and measurement.

Detailed site descriptions and additional methods and results in the Panamanian island ecosystem can be found in ref. 26. All animals were handled and treated ethically within the guidelines of the Institutional Animal Care and Use Committee.

Landscape habitat classifications. To characterize landscape context in Costa Rica we conducted a manual habitat classification by digitizing aerial photographs. Manual habitat classification was favoured over automated approaches to facilitate the accurate identification of fine-scale landscape features in a complex countryside ecosystem³². Orthorectified aerial photographs used in the habitat classification were from Costa Rica Airborne Research and Technology Applications taken in 2003 and 2005 at 2 m spatial resolution. The habitat map covered 23,600 ha and included all sampling locations. The area was delineated to encompass all study locations sampled by Stanford University's Center for Conservation Biology (see, for example, refs 33, 42–48). The four ecosystem elements digitized at 2 m resolution over this extent included forest elements (primary and secondary forest of all sizes, single trees, *charral* (early secondary forest), large shrubs, live fences, hedgerows, non-native timber and fruit tree plantations, and non-native garden ornamentals; Fig. 1); agricultural plots (the majority being sun coffee plantations, but also including banana, plantain, tomato, and chili plantations); pasture; and houses, roads and other rural infrastructure. The final map was verified in the field from multiple vantage points, compared with previous studies in the landscape and checked with 150 random samples across the landscape. Forest elements from the manual habitat classification were used in analyses, whereas agricultural plots, pasture, houses, roads and other rural infrastructure were broadly treated as deforested habitat.

To characterize landscape context in Panama we used the 'BCI Basemap' data set⁴⁹ to delineate shorelines and islands of Lake Gatún with the 'BCI Landcover Map 2003' data set⁵⁰ to confirm forest, water and agriculture cover. Both data sets were accessed in early 2013 online through the Smithsonian Tropical Research Institute GIS Data Portal (<http://mapserver.stri.si.edu/geonetwork/>). All islands were completely forested except for 1 ha on a 16-ha island.

Landscape explanatory variable calculations. Several relevant landscape metrics were calculated for the Costa Rican countryside ecosystem and the Panamanian island ecosystem. The landscape metrics included measures that captured the related variables of habitat area and habitat isolation. We also examined the importance of spatial scale when measuring some of the landscape characteristics in this study. The specific landscape variables calculated and used in statistical modelling included fragment/island size, isolation from contiguous tropical forest, general habitat type, and total forest area within a range of spatial scales. We do not report effects of forest or island fragmentation in this study because measures of fragmentation are often confounded with area and merit a separate, in-depth analysis for clarity⁵¹.

First, we calculated forest fragment or island size in each ecosystem. In the Panamanian island ecosystem each island was clearly distinct and the contiguous forest habitat on each island was easily measured as the total size of the island. The forest fragments in the countryside ecosystem of Costa Rica, however, were slightly more difficult to measure: forest fragments are not easily distinguished because of many fine-scale forest connections across the farming landscape³³ (Fig. 2). In the countryside ecosystem nearly every forest fragment is connected to other forest remnants by live fences, riparian strips and fine-scale ribbons and clusters of trees that weave across ridges, valleys and property boundaries. In most cases these fine-scale forest connections are less than 20 m in width. We therefore arbitrarily omitted forest element connections that were less than 20 m in width to calculate forest fragment size in the countryside ecosystem.

Second, habitat isolation for the countryside and island ecosystems was measured as the Euclidean distance from the centre of a site or mist net to the nearest edge of the Las Cruces Biological Reserve, La Amistad International Park or the nearest Panamanian mainland shore. Sites located inside reserves or mainland peninsulas received negative values that measured their distances from reserve edges and mainland shores.

Third, we included general and broad habitat types as discrete explanatory variables in our models. In the countryside ecosystem each site and net was labelled as reserve, forest fragment or coffee plantation. In the island ecosystem we assigned sites and nets into three categories: mainland interior, mainland edge or island, following ref. 26. Moreover, some categories were lumped during analysis to check for broader differences between islands and all mainland sites in the Panamanian island ecosystem or between forest and non-forest sites in the Costa Rican countryside ecosystem.

Finally, we calculated the total forest area—contiguous or not—for each site and each mist net at multiple spatial scales. For example, the total amount of forest area at each site and mist net, including nearby forest fragments and neighbouring islands, was measured at radii ranging from 50 m to 1 km (measurements were calculated at 50-m increments).

Selecting spatial scales for measuring forest area. Countryside ecosystems and island ecosystems differ substantially because of landscape context. Countryside ecosystems often contain high proportions of human-made habitat that, generally, includes considerable vegetative structure and, depending on the spatial scales used to view habitat characteristics, many smaller 'patches' of habitat—often embodied in the form of single trees or fallowed agricultural fields. The surrounding landscape

context in island ecosystems is strictly homogeneous, except for neighbouring islands and the mainland shore. In view of the substantial differences in landscape context and recently proposed ideas, specifically the habitat amount hypothesis¹¹, we champion total habitat area within an appropriately selected spatial scale as a more ecologically relevant measure of habitat.

We used an established approach described in detail in ref. 33 to compare and select appropriate spatial scales during analysis. All results were robust to forest area measured at different spatial scales. For clarity, two spatial scales were selected to measure forest area after *a priori* and *post hoc* analyses, and statistical modelling revealed that few differences in results were observed when forest area was measured among a broad range of spatial scales. We chose spatial scales that emerged as strong predictors after we verified that forest area measured at all spatial scales generally agreed. Site-level analyses used forest area measured within a 1-km spatial scale, and net-level analyses used forest area measured within a 150-m spatial scale. We do not report the nuances observed between different spatial scales in this study, which merit a separate, in-depth analysis to determine their ecological relevance¹¹.

Dimensions of bat biodiversity: species richness estimates. Bat captures in the countryside and island ecosystems were used to inform three dimensions of biodiversity metrics at the resolution of a site or a mist net. The first dimension of bat biodiversity was an estimate of bat species richness at each site (that is, the estimated number of unique species at each site). Chao species richness estimates of bats were used because they perform well with small sample sizes and uneven sampling effort⁵². Chao species richness estimates were used in analyses except when the standard error surrounding the species richness estimate was undefined when calculated in R v.2.15.2 with the *vegan* package. In cases in which species richness estimates were obtuse, the observed species richness (that is, the observed number of unique species captured at a given site) was used in analysis. In all, four sites used observed species richness values (two Costa Rican coffee plantation sites, one Costa Rican forest fragment site and one Panamanian island site). Results were consistent when observed species richness was used, but the inclusion of Chao species richness estimates dealt with differences in sampling effort between studies in Costa Rica and Panama.

Dimensions of bat biodiversity: species evenness estimates. The second dimension of bat biodiversity is Pielou's species evenness index; this was calculated for each site. Species evenness refers to how equal the assemblage of bat species is numerically, or how close in abundance species at a site are. Pielou's species evenness index is based on the Shannon–Weaver diversity index (denoted by H'):

$$H' = - \sum_{i=1}^S p_i \ln p_i \quad (1)$$

In the Shannon–Weaver diversity index, p_i is the proportion of species i at a site, and S is the number of species at a site. Pielou's species evenness index, J , was then calculated for each site; it is defined as

$$J = \frac{H'}{\ln S} \quad (2)$$

Dimensions of bat biodiversity: Assemblage Abundance Shift Index. The third dimension of bat biodiversity in this study centred on changes in abundance-based assemblages of bats relative to bat assemblages found in the baseline habitats for each ecosystem. Bat assemblage abundance shifting was calculated for each mist net because field observations and preliminary analyses of bat assemblages in the countryside ecosystem suggested that differences in the bat assemblage existed between nets within the same site relative to a net's proximity to nearby trees, coffee plantations and pastures. For example, patterns of assemblage differences emerged nightly at a given site when field assistants compared species and abundances of bats extracted from groups of nets closer to habitat transitions with those of bats captured in nets located uniformly in forested habitat or coffee plantations.

The degree of assemblage abundance shift was based on an index we developed that compared a net's abundance-based assemblage similarity to that of the set of nets located in the Las Cruces Biological Reserve and La Amistad International Park, in the countryside ecosystem, or the Panamanian mainland, in the island ecosystem. We used Chao abundance-based similarity coefficients to calculate the similarity of a net's abundance-based assemblage to that of all the nets in each ecosystem baseline. The Chao abundance-based similarity coefficient measures similarity between two samples, and implements a Jaccard-type index that accounts for unseen, shared species; it is defined as

$$\hat{J}abd(U, W) = \frac{\hat{U}\hat{W}}{\hat{U} + \hat{W} - \hat{U}\hat{W}} \quad (3)$$

In equation (3), \hat{U} and \hat{W} represent estimates of each site in such a way that they are adjusted for unseen shared species. To account for the effect of unseen shared

species, the index uses frequencies of observed, rare, shared species to estimate an appropriate adjustment term that accounts for the unseen, shared species. The adjustment is achieved by using the following equations for each estimate:

$$\hat{U} = \sum_{i=1}^{D_{12}} \frac{X_i}{n} + \frac{(m-1)}{m} \frac{f_{+1}}{2f_{+2}} \sum_{i=1}^{D_{12}} \frac{X_i}{n} I(Y_i = 1) \quad (4)$$

$$\hat{W} = \sum_{i=1}^{D_{12}} \frac{Y_i}{m} + \frac{(n-1)}{n} \frac{f_{+1}}{2f_{+2}} \sum_{i=1}^{D_{12}} \frac{Y_i}{m} I(X_i = 1) \quad (5)$$

In equations (4) and (5) each function accounts for unseen, shared species between two sites by using the following logic: if a random sample of n individuals (sample 1) is taken from site 1 and a random sample of m individuals (sample 2) from site 2, and species frequencies in each sample are denoted as $(X_1, X_2, \dots, X_{S_{12}})$ and $(Y_1, Y_2, \dots, Y_{S_{12}})$, respectively, with zero used where species were missing from the sample. Then the pairs of frequencies for the S_{12} species truly shared by the two assemblages are $(X_1, Y_1)(X_2, Y_2) \dots (X_{S_{12}}, Y_{S_{12}})$. Assume that D_{12} of the S_{12} shared species available are actually observed in both samples and that their frequencies are the first D_{12} pairs. Thus, an additional $S_{12} - D_{12}$ species are shared by the two assemblages but absent from one or both samples.

In equations (4) and (5) the indicator function $I(\text{expression})$ is defined such that $I = 1$ if the expression is true, and $I = 0$ if false. Let $f_{1+} = \sum_{i=1}^{D_{12}} I[X_i = 1, Y_i \geq 1]$ be the observed number of shared species that are singletons ($X_i = 1$) in sample 1 (these species must be present in sample 2, but may have any abundance). Now, let f_{2+} represent the observed number of shared species that are doubletons ($X_i = 2$) in sample 1. Similarly, f_{+1} and f_{+2} are the observed number of shared species that are, respectively, singletons ($Y_i = 1$) and doubletons ($Y_i = 2$) in sample 2. If $f_{+2} = 0$ or $f_{2+} = 0$, replace f_{+2} and f_{2+} in the denominators by $f_{+2} + 1$ or $f_{2+} + 1$, respectively. If \hat{U} or \hat{W} is greater than 1, then it is replaced by 1. See ref. 53 for more details, explanations and derivation of equations.

The Chao abundance-based similarity coefficients were calculated for each net (denoted as k) with all nets in the Las Cruces Biological Reserve and La Amistad International Park in the countryside ecosystem, or the mainland in the Panamanian island ecosystem (denoted as b_i), using variations of equations (3) and (4) in R v.2.15.2 with the *vegan* package. The arithmetic means of all Chao abundance-based similarity coefficients for each net with each baseline sample were calculated (denoted as b):

$$M_b = \frac{1}{n} \cdot \sum_{i=1}^n \hat{J}abd(k, b_i) \quad (6)$$

The mean Chao abundance-based similarity coefficients for each baseline were also calculated relative to all baseline nets for each ecosystem. Finally, we standardized all values by dividing each M_b value by the maximum M_b value in each ecosystem ($M_b \text{max}$):

$$M = \frac{M_b}{M_b \text{max}} \quad (7)$$

The resulting values for each net (that is, M) are between 0 and 1, and represent the abundance-based assemblage similarity of a single net to all the nets located in the corresponding ecosystem baseline, while accounting for detection biases of unseen, shared species. Therefore, the closer a net's value is to 1 in the countryside ecosystem, the more similar that net is in bat species composition and abundances to those of all the nets located inside the Las Cruces Biological Reserve and La Amistad International Park. Similarly, the closer a net's value is to 0 in the island ecosystem, the less similar that net is in bat species composition and abundances to those of all the nets located on the Panamanian mainland.

As calculated in equations (3)–(7), the Abundance-Based Assemblage Similarity Index (that is, M) is dependent on changes in species richness and abundances. The Abundance-Based Assemblage Similarity Index was therefore modified to account for species richness. To account for the effect of species richness in the Abundance-Based Assemblage Similarity Index we performed a regression between that index and the observed species richness values for each net and calculated the residuals from the best-fit model (Extended Data Fig. 2). Nets with five or fewer total captures were excluded from the Assemblage Shift Index (21 nets from the countryside ecosystem and zero nets from the island ecosystem). Positive residuals indicate a greater similarity than expected in assemblage abundances of a net to those of all nets in the corresponding baseline. Negative residuals indicate greater differences in the assemblage abundances than expected.

In its final form, the Assemblage Abundance Shift Index was independent of observed species richness, estimated species richness, species evenness, and total bat abundance. The Assemblage Abundance Shift Index is a holistic measure of how the bat assemblage changes in abundances across space, independently of changes in species richness.

Statistical analyses. All statistical analyses were conducted in a GLMM (generalized linear mixed effects models) framework⁵⁴. Analyses were conducted in R v.2.15.2 (R Development Core Team 2012) and relied extensively on vegan, lme4, nlme and AICcmodavg packages. All variables were tested for normality to meet assumptions of statistical tests.

In our first study contrasting a Costa Rican countryside ecosystem and an island ecosystem, we conducted three suites of analyses for each dimension of bat biodiversity (that is, species richness estimates, species evenness estimates, and the Assemblage Abundance Shift Index). Each set of analyses was designed to identify the effects of ecosystem change on bat biodiversity—specifically concentrating on fragment/island size, isolation from contiguous tropical forest, general habitat type, and total forest area within a wide range of spatial scales (see the section above on selecting spatial scales for measuring forest area). All models avoided collinearity of explanatory variables by disqualifying any linear models that used multiple dependent variables (for example, fragment or island size and total forest area are collinear and were therefore not used in the same statistical model to explain any of the bat biodiversity response variables).

At the onset of each suite of analyses we explicitly tested for interaction terms between landscape variables and the ecosystem type (that is, countryside ecosystem versus island ecosystem). The logic behind testing for interaction terms stems from the fact that the countryside ecosystem and the island ecosystem are separated by about 350 km; they have slightly different species pools and therefore different intercepts within a linear modelling framework. We therefore tested statistically for differences between ecosystems through differences in slopes of functions by using interaction terms in our modelling, rather than only looking for differences in the intercept value.

In each suite of analyses we first tested for differences between the countryside ecosystem and the island ecosystem and restricted our analysis to forested habitats by excluding data from coffee plantations (Extended Data Tables 1 and 2). After differences were confirmed between the two ecosystem types, the second suite of analyses included data from coffee plantations and focused on fitting the best model and selecting the best landscape variable or variables for predicting the bat biodiversity response variable of interest (for example estimated species richness, estimated species evenness or the Assemblage Abundance Shift Index).

After differences between ecosystems were tested, each ecosystem was modelled independently (that is, analyses were performed with data from only a single ecosystem, rather than including a factor within each model that denoted ecosystem type as countryside ecosystem or island ecosystem). Nonlinear functions were included in ecosystem-specific analyses, and the most relevant models are presented in Extended Data Tables 3 and 4.

In all analyses, models were compared by using the corrected Akaike information criterion (AIC_c) and maximum likelihood. Because landscape forest areas measured at multiple spatial scales were autocorrelated and collinear, we identified a single appropriate spatial scale by comparing models using AIC_c and maximum likelihood. Little differences in AIC_c values were observed between landscape variables measured at different spatial scales.

In the GLMM framework we made use of spatial correlation structures to avoid issues of spatial autocorrelation⁵¹. The UTM (Universal Transverse Mercator) coordinates of each site or net, measured as the centre of each site or net, were used to account for spatial autocorrelation between sites and nets. Models were compared that used a variety of correlation structures, including 'corSpher', 'corLin', 'corRatio', 'corGaus' and 'corExp' (see pp. 161–170 of ref. 54 for more information). The 'corRatio' correlation structure, or the ratio correlation structure, consistently performed best in all models and in all analyses. All analyses make use of the ratio correlation structure with a 'nugget'. Moreover, site identity was included as a fixed, random effect in analyses conducted at the net level.

The three suites of analyses focused on identifying how the response variables of, respectively, bat species richness estimates, bat species evenness estimates, and the bat Assemblage Abundance Shift Index respond to fragment or island size, isolation from contiguous tropical forest, general habitat type, and total forest area within a range of spatial scales. The model comparisons for each suite of analyses are presented in Extended Data Tables 1–4. Results from models that included forest area measured at multiple spatial scales are not presented for clarity, but few differences between models were observed (see the section above on selecting spatial scales for measuring forest area).

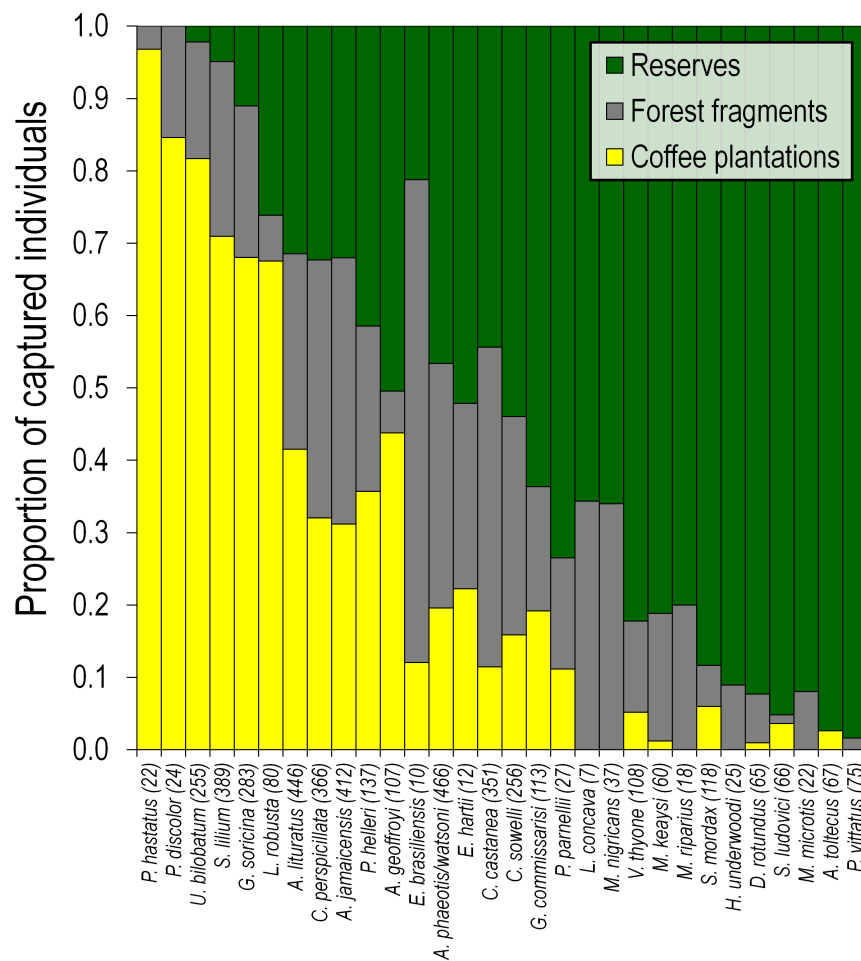
Meta-analysis of studies of bat biodiversity across countryside and island ecosystems worldwide. We searched for all relevant research articles from any time period by using the following Web of Science search query: Title = (bat) AND

Topic = (conservation and fragment*), Timespan = All years, Search language = English. Moreover, we asked authors of many of the research articles we reviewed to ensure that our list was representative—especially authors whose studies were conducted outside the Neotropics (Fig. 4a). From this list we reviewed articles and retained studies that included comparisons between habitats, studies that sampled a minimum of ten different bat species, and studies with replicate sampling sites and similar sampling effort in each habitat type.

For studies that met our criteria, we compared species richness between habitat types in each study. Habitat types in each study fell into the following three broad categories: minimally altered reference habitats (for example forest reserves or contiguous mainland forest); fragmented or isolated reference habitat (for example forest fragments and natural or human-made, true islands); or human-made or naturally open habitat (for example cropland, pasture and, in one study, African savanna). When sufficient data were available, the bias-corrected Hedge's g^* was calculated and used to determine whether bat species richness differed between habitat types. Reported statistical differences in species richness were used in studies if there were insufficient data to calculate Hedge's g^* .

Wide arrays of metrics were used across studies to report bat assemblage shifts between habitat types. Because it is difficult to compare multiple measures of bat assemblage shift between studies, we chose to report assemblage shifts in each study as observed or unobserved. The detailed list of studies used is presented in Supplementary Data.

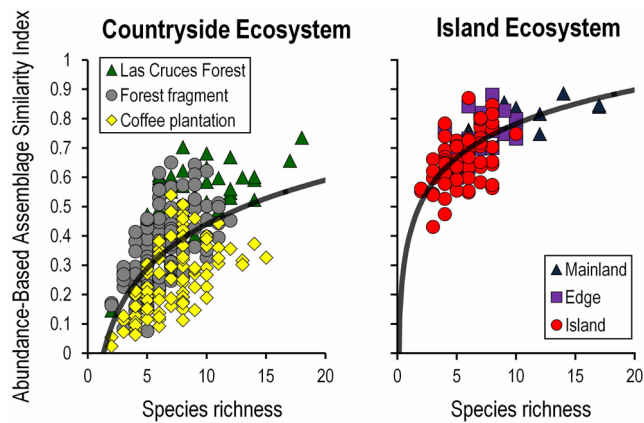
31. Stiles, F. & Skutch, A. *A Guide to the Birds of Costa Rica* 632 (Instituto Nacional de Biodiversidad, 1989).
32. Holdridge, L. R. *Life Zone Ecology* 146 (Tropical Science Center, 1979).
33. Mendenhall, C., Sekercioglu, C., Oviedo Brenes, F., Ehrlich, P. & Daily, G. Predictive model for sustaining biodiversity in tropical countryside. *Proc. Natl Acad. Sci. USA* **108**, 16313–16316 (2011).
34. Sansonetti, V. *Quemé Mis Naves en estas Montañas: La Colonización de la Altiplanicie de Coto Brus y la Fundación de San Vito de Java* 99 (Jiménez & Tanzi, 1995).
35. Clement, R. & Horn, S. Pre-Columbian land-use history in Costa Rica. *Holocene* **11**, 419–426 (2001).
36. Mann, C. 1491: *New Revelations of the Americas before Columbus* 541 (Knopf, 2005).
37. Kunz, T. & Parsons, P. *Ecological and Behavioral Methods for the Study of Bats* 920 (Johns Hopkins Univ. Press, 2011).
38. Leigh, E. Jr, Rand, S. & Windsor, D. *The Ecology of a Tropical Forest: Seasonal Rhythms and Long-Term changes* 480 (Smithsonian Institution Press, 1982).
39. Bonaccorso, F., Smythe, N. & Humphrey, S. Improved techniques for marking bats. *J. Mamm.* **57**, 181–182 (1976).
40. Leigh, E. Jr, Wright, S., Herre, E. & Putz, F. The decline of tree diversity on newly isolated tropical islands. *Evol. Ecol.* **7**, 76–102 (1993).
41. Handley, C. Jr, Wilson, D. & Gardner, A. *Demography and Natural History of the Common Fruit Bat, *Artibeus jamaicensis*, on Barro Colorado Island, Panama* 173 (Smithsonian Institution Press, 1991).
42. Mendenhall, C., Daily, G. & Ehrlich, P. Improving estimates of biodiversity loss. *Biol. Conserv.* **151**, 32–34 (2012).
43. Daily, G. C., Ehrlich, P. R. & Sánchez-Azofeifa, G. A. Countryside biogeography: use of human-dominated habitats by the avifauna of southern Costa Rica. *Ecol. Appl.* **11**, 1–13 (2001).
44. Ricketts, T., Daily, G., Ehrlich, P. & Fay, J. Countryside biogeography of moths in a fragmented landscape. *Conserv. Biol.* **15**, 378–388 (2001).
45. Horner-Devine, M., Daily, G., Ehrlich, P. & Boggs, C. Countryside biogeography of tropical butterflies. *Conserv. Biol.* **17**, 168–177 (2003).
46. Mayfield, M. & Daily, G. Countryside biogeography of Neotropical herbaceous and shrubby plants. *Ecol. Appl.* **15**, 423–439 (2005).
47. Brosi, B., Daily, G., Shih, T., Oviedo Brenes, F. & Duran, G. The effects of forest fragmentation on bee communities in tropical countryside. *J. Appl. Ecol.* **45**, 773–783 (2008).
48. Karp, D. et al. Forest bolsters bird abundance, pest control and coffee yield. *Ecol. Lett.* **16**, 1339–1347 (2013).
49. Smithsonian Tropical Research Institute GIS Section. BCI Basemap. Scale 1:50 000 (Smithsonian, 2012).
50. Panama Canal Authority Remote Sensing Unit. BCI Landcover Map 2003 (Cobertura Boscosa 2003—Área de BCI). (Panama Canal Authority, 2003).
51. Fahrig, L. Rethinking patch size and isolation effects. *J. Biogeogr.* **40**, 1649–1663 (2013).
52. Chao, A. Estimating the population size for capture-recapture data with unequal catchability. *Biometrics* **43**, 783–791 (1987).
53. Chao, A., Chazdon, R., Colwell, R. & Shen, T. A new statistical approach for assessing similarity of species composition with incidence and abundance data. *Ecol. Lett.* **8**, 148–159 (2005).
54. Zuur, A., Ieno, E., Walker, N., Saveliev, A. & Smith, G. *Mixed Effects Models and Extensions in Ecology with R* 574 (Springer, 2009).



Bats species with ≥ 7 individuals captured
(ordered by forest dependency rank)

Extended Data Figure 1 | Bats use a variety of habitats in the countryside ecosystem. Shown are the proportions of captured individuals from 30 species in different countryside habitats. Forest dependence rank ranges from forest avoidance (left side of *x* axis) to reserve dependent (right side of *x* axis) and was determined by comparing relative abundance in reserves (green) with that in

coffee plantations (yellow). The proportions of individuals captured in smaller forest fragments of various sizes are also shown. Total numbers of individuals per species are listed parenthetically after abbreviated species names. A total of 4,424 individuals are represented.



Extended Data Figure 2 | Assemblage Abundance Shift Index based on ordination analyses of bat abundances and how they collectively shift relative to bat abundances in minimally altered habitat. The plots demonstrate how the Assemblage Abundance Shift Index accounted for changes in species richness to focus on predicting changes in assemblage-level shifts in abundances between habitats. In both ecosystems regression analyses favoured logarithmic relationships between the abundance-based assemblage similarity of the bats captured in a net relative to the reserve or mainland nets and the observed species richness of the bats captured in the net (see Methods). Logarithmic models (solid lines) outperformed linear models in model comparisons (countryside ecosystem $\Delta AIC_c = 22.75$; island ecosystem $\Delta AIC_c = 5.92$). For each ecosystem, logarithmic models were used to calculate the residual assemblage shift for each net that was not explained by changes in species richness but by changes in the abundances of species. The residuals are therefore an index of assemblage abundance shifting after accounting for changes in species richness. Regression coefficients and statistics are described in Extended Data Table 3.

Extended Data Table 1 | Summary of model performances of forest habitat comparisons in an island ecosystem and a countryside ecosystem

Model structure for predicting: Estimated species richness (Chao 1)						
total forest area within 1 km [%] * ecosystem type [factor]	7	176.58	0.00	0.59	0.59	-78.75
general habitat type [factors: island, mainland edge, mainland interior, forest fragment, reserve]	8	178.78	2.20	0.20	0.79	-77.96
log(fragment/island size [ha]) + isolation distance [km] * ecosystem type [factor]	8	180.56	3.98	0.08	0.87	-78.85
total forest area within 1 km [%] + ecosystem type [factor]	6	181.33	4.75	0.06	0.93	-82.84
log(fragment/island size [ha]) * ecosystem type [factor]	7	181.84	5.26	0.04	0.97	-81.38
total forest area within 1 km [%]	5	184.80	8.22	0.01	0.98	-86.15
log(fragment/island size [ha]) + ecosystem type [factor]	6	185.14	8.55	0.01	0.99	-84.74
log(fragment/island size [ha]) + isolation distance [km] + ecosystem type [factor]	7	186.65	10.07	0.00	0.99	-83.78
isolation distance [km] * ecosystem type [factor]	7	187.11	10.53	0.00	1.00	-84.01
isolation distance [km] + ecosystem type [factor]	6	187.48	10.90	0.00	1.00	-85.92
log(fragment/island size [ha])	5	188.58	12.00	0.00	1.00	-88.04
log(fragment/island size [ha]) + isolation distance [km]	6	191.03	14.45	0.00	1.00	-87.69
Null [intercept]	4	195.35	18.77	0.00	1.00	-92.88
isolation distance [km]	5	195.73	19.14	0.00	1.00	-91.61

Model structure for predicting: Estimated species evenness (Pielou)						
isolation distance [km] * ecosystem type [factor]	7	-36.47	0.00	0.97	0.97	27.78
isolation distance [km] + ecosystem type [factor]	6	-28.21	8.26	0.02	0.99	21.93
total forest area within 1 km [%] + ecosystem type [factor]	6	-26.04	10.43	0.01	0.99	20.85
log(fragment/island size [ha]) + isolation distance [km] + ecosystem type [factor]	7	-24.91	11.56	0.00	1.00	22.00
Null [intercept]	4	-22.19	14.27	0.00	1.00	15.90
general habitat type [factors: island, mainland edge, mainland interior, forest fragment, reserve]	8	-22.09	14.37	0.00	1.00	22.48
log(fragment/island size [ha]) + ecosystem type [factor]	6	-19.61	16.85	0.00	1.00	17.63
total forest area within 1 km [%] * ecosystem type [factor]	7	-19.51	16.96	0.00	1.00	19.30
log(fragment/island size [ha])	5	-19.43	17.04	0.00	1.00	15.97
total forest area within 1 km [%]	5	-19.36	17.11	0.00	1.00	15.93
log(fragment/island size [ha]) + isolation distance [km]	6	-18.31	18.15	0.00	1.00	16.98
log(fragment/island size [ha]) + isolation distance [km] * ecosystem type [factor]	8	-16.96	19.50	0.00	1.00	19.91
log(fragment/island size [ha]) * ecosystem type [factor]	7	-16.74	19.73	0.00	1.00	17.91
isolation distance [km]	5	6.67	43.13	0.00	1.00	2.92

Model structure for predicting: Assemblage Abundance Shift Index						
total forest area within 150 m [%] + ecosystem type [factor]	7	-822.21	0.00	0.61	0.61	418.30
log(fragment/island size [ha]) * ecosystem type [factor]	8	-819.93	2.28	0.20	0.81	418.22
total forest area within 150 m [%]	6	-819.25	2.96	0.14	0.95	415.77
log(fragment/island size [ha]) + ecosystem type [factor]	7	-814.46	7.75	0.01	0.96	414.43
total forest area within 150 m [%] * ecosystem type [factor]	8	-814.44	7.77	0.01	0.98	415.47
log(fragment/island size [ha]) + isolation distance [km] + ecosystem type [factor]	8	-814.03	8.18	0.01	0.99	415.27
log(fragment/island size [ha]) + isolation distance [km] * ecosystem type [factor]	9	-812.19	10.02	0.00	0.99	415.41
log(fragment/island size [ha])	6	-811.89	10.32	0.00	0.99	412.09
general habitat type [factors: island, mainland edge, mainland interior, forest fragment, reserve]	9	-810.56	11.65	0.00	1.00	414.60
isolation distance [km] + ecosystem type [factor]	7	-810.36	11.85	0.00	1.00	412.38
log(fragment/island size [ha]) + isolation distance [km]	7	-809.93	12.28	0.00	1.00	412.16
isolation distance [km] * ecosystem type [factor]	8	-808.51	13.70	0.00	1.00	412.51
Null [intercept]	5	-805.75	16.46	0.00	1.00	407.98
isolation distance [km]	6	-805.35	16.86	0.00	1.00	408.82

The summary shows comparisons of forest fragments embedded in a human-dominated landscape and true forest islands surrounded by water; see Extended Data Table 3 for analyses with deforested habitat matrix sampling included. Listed are relevant outputs summarizing the performances of generalized linear mixed effects models predicting estimated bat species richness (green heading), bat species evenness (orange heading) and the Species Assemblage Abundance Shift Index (purple heading; see Methods), using a variety of landscape metrics as explanatory variables. Bat biodiversity and landscape data were gathered from a Costa Rican countryside ecosystem and a Panamanian island ecosystem (that is, ecosystem type). All models are generalized linear mixed-effects models with Gaussian error structures and identity links. In addition, all models include a 'corRatio' correlation structure based on site or net coordinates (UTM). Net-level analyses included site identity as a random effect in all models.

Extended Data Table 2 | Regression coefficients and relevant statistics generated from best-fit models from Extended Data Table 1

Estimated species richness (Chao1)	Estimate	SE	DF	t-value	p-value
(intercept)	8.923	1.476	30	6.048	0.000
total forest area within 1 km [%]	18.818	3.602	30	5.225	0.000
ecosystem type [factor]	15.411	3.141	30	4.906	0.000
total forest area within 1 km * ecosystem type	-18.236	6.222	30	-2.931	0.007

Estimated species evenness (Pielou)	Estimate	SE	DF	t-value	p-value
(intercept)	0.575	0.036	30	15.948	0.000
isolation distance [km]	-0.114	0.024	30	-4.672	0.000
ecosystem type [factor]	0.265	0.060	30	4.395	0.000
isolation distance * ecosystem type	0.107	0.027	30	3.960	0.001

Assemblage Abundance Shift Index	Estimate	SE	DF	t-value	p-value
(Intercept)	-0.052	0.020	262	-2.641	0.009
total forest area within 150 m [%]	0.074	0.025	262	2.979	0.003
ecosystem type [factor]	0.029	0.013	28	2.279	0.031

The models explain relationships between three dimensions of bat biodiversity and the remaining landscape variables after model optimization for forest habitats in an island ecosystem and a countryside ecosystem. Model performances and comparisons are listed in Extended Data Table 1. Summarized models are in bold in Extended Data Table 1. After model selection, best-fit models were rerun and are presented here using restricted maximum likelihood³⁴. Data from the deforested habitat in the countryside ecosystem (that is, sun coffee plantations) are not included in these analyses. In-depth ecosystem-specific analyses that include the deforested habitat in the countryside ecosystem are presented in Extended Data Tables 4 and 5.

Extended Data Table 3 | Summary of model performances of ecosystem-specific models

Model structure for predicting: Estimated species richness (Chao 1)						
	K	AICc	ΔAICc	AICcW	CumW	LL
Countryside ecosystem						
broad habitat type [factors: coffee plantation, forest]	5	113.93	0.00	0.49	0.49	-49.66
general habitat type [factors: coffee plantation, forest fragment, reserve]	6	114.15	0.22	0.44	0.94	-47.58
total forest area within 1 km [%]	5	120.76	6.83	0.02	0.95	-53.07
isolation distance [km]	5	120.86	6.92	0.02	0.97	-53.12
Null [intercept]	4	121.03	7.09	0.01	0.98	-55.08
log(fragment size [ha])	5	121.10	7.16	0.01	1.00	-53.24
log(fragment size [ha]) + isolation distance [km]	6	123.92	9.99	0.00	1.00	-52.46
Island ecosystem						
total forest area within 1 km [%]	5	100.61	0.00	0.87	0.87	-42.58
broad habitat type [factors: coffee plantation, forest]	5	106.32	5.71	0.05	0.92	-45.43
log(island size [ha])	5	106.70	6.09	0.04	0.97	-45.62
general habitat type [factors: coffee plantation, forest fragment, reserve]	6	107.77	7.17	0.02	0.99	-43.69
log(island size [ha]) + isolation distance [km]	6	109.91	9.31	0.01	1.00	-44.76
isolation distance [km]	5	113.26	12.65	0.00	1.00	-48.90
Null [intercept]	4	114.67	14.06	0.00	1.00	-51.67
Model structure for predicting: Estimated species evenness (Pielou)						
	K	AICc	ΔAICc	AICcW	CumW	LL
Countryside ecosystem						
Null [intercept]	4	-53.24	0.00	0.56	0.56	32.05
isolation distance [km]	5	-50.19	3.05	0.12	0.69	32.40
broad habitat type [factors: coffee plantation, forest]	5	-49.59	3.65	0.09	0.78	32.10
total forest area within 1 km [%]	5	-49.49	3.75	0.09	0.86	32.05
log(fragment size [ha])	5	-49.49	3.75	0.09	0.95	32.05
general habitat type [factors: coffee plantation, forest fragment, reserve]	6	-47.58	5.66	0.03	0.98	33.29
log(fragment size [ha]) + isolation distance [km]	6	-46.12	7.12	0.02	1.00	32.56
Island ecosystem						
isolation distance [km]	5	-7.99	0.00	0.77	0.77	11.72
log(island size [ha]) + isolation distance [km]	6	-4.71	3.27	0.15	0.92	12.56
Null [intercept]	4	-3.24	4.74	0.07	0.99	7.29
total forest area within 1 km [%]	5	2.25	10.24	0.00	0.99	6.60
general habitat type [factors: coffee plantation, forest fragment, reserve]	6	3.01	11.00	0.00	1.00	8.70
broad habitat type [factors: coffee plantation, forest]	5	3.38	11.37	0.00	1.00	6.04
log(island size [ha])	5	3.91	11.89	0.00	1.00	5.77
Model structure for predicting: Assemblage Abundance Shift Index						
	K	AICc	ΔAICc	AICcW	CumW	LL
Countryside ecosystem						
ln(total forest within 150 m [%])	6	-874.88	0.00	0.91	0.91	443.59
total forest area within 150 m [%]	6	-870.21	4.67	0.09	1.00	441.26
general habitat type [factors: coffee plantation, forest fragment, reserve]	7	-852.71	22.17	0.00	1.00	433.56
broad habitat type [factors: coffee plantation, forest]	6	-851.33	23.55	0.00	1.00	431.82
log(fragment size [ha])	6	-847.65	27.23	0.00	1.00	429.98
log(fragment size [ha]) + isolation distance [km]	7	-846.02	28.86	0.00	1.00	430.22
isolation distance [km]	6	-838.98	35.90	0.00	1.00	425.64
Null [intercept]	4	-671.68	203.20	1.00	1.00	339.91
Island ecosystem						
log(island size [ha])	6	-241.10	0.00	0.37	0.37	126.99
broad habitat type [factors: island, mainland]	6	-240.48	0.62	0.27	0.64	126.68
total forest area within 350 m [%]	6	-239.24	1.86	0.14	0.78	126.06
log(island size [ha]) + isolation distance [km]	7	-238.82	2.28	0.12	0.90	127.01
general habitat type [factors: island, mainland edge, mainland interior]	7	-238.17	2.93	0.08	0.98	126.68
isolation distance [km]	6	-233.82	7.28	0.01	0.99	123.35
Null [intercept]	5	-233.53	7.57	0.01	1.00	122.08

The models used were those predicting estimated bat species richness (green heading), bat species evenness (orange heading) and the Assemblage Abundance Shift Index (purple heading), using a series of landscape metrics as explanatory variables. All models are generalized linear mixed-effects models with Gaussian error structures and identity links. In addition, all models include a 'corRatio' correlation structure based on each site's or net's UTM coordinates. Net-level analyses included site identity as a random effect in all models.

Extended Data Table 4 | Regression coefficients and relevant statistics generated from best-fit models from Extended Data Table 3

Estimated species richness (Chao1)	Estimate	SE	t-value	p-value
Countryside ecosystem				
(Intercept)	18.389	1.425	12.900	0.000
broad habitat type (forest) [factors: coffee plantation, forest]	6.235	1.723	3.618	0.002
Island ecosystem				
(Intercept)	18.488	1.351	13.682	0
habitat type (forest fragment) [factors: coffee plantation, forest fragment, reserve]	4.997	1.698	2.943	0.01
habitat type (reserve) [factors: coffee plantation, forest fragment, reserve]	9.016	2.062	4.373	0.001
Island ecosystem				
(Intercept)	8.953	1.309	6.837	0.000
total forest area within 1 km [%]	18.680	3.202	5.834	0.000
Island ecosystem				
(Intercept)	9.814	1.490	6.586	0.000
log(island size [ha])	2.623	0.572	4.582	0.000
Island ecosystem				
(Intercept)	17.304	1.592	10.87	0.000
isolation distance [km]	-2.612	1.081	-2.417	0.029
Estimated species evenness (Pielou)	Estimate	SE	t-value	p-value
Countryside ecosystem				
(Intercept)	0.818	0.011	76.064	0.000
Island ecosystem				
(Intercept)	0.575	0.045	12.758	0.000
isolation distance [km]	-0.114	0.031	-3.738	0.002
Assemblage Abundance Shift Index	Estimate	SE	t-value	p-value
Countryside ecosystem				
(Intercept)	0.055	0.011	5.105	0.000
ln(total forest area within 150 m [%])	0.063	0.009	7.366	0.000
Island ecosystem*				
(Intercept)	0.030	0.012	2.557	0.012
broad habitat type (island) [factors: island, mainland]	-0.046	0.015	-3.169	0.006

*The second most competitive model is presented for predicting the Assemblage Abundance Shift Index in the island ecosystem. Island size and broad habitat types in the island ecosystem (i.e., islands and mainland) are confounded. Moreover, island size did not explain the community shift index when mainland sites were excluded from analysis (regression statistics for islands only: $t = 0.561$, $P = 0.588$, $N = 66$). Therefore, we concluded that island size is not a better predictor of community shift than broad habitat type. Model coefficients and statistics correspond with the best-fit lines in panel **c** of Figure 3 in the main text.

The models explain ecosystem-specific relationships between three dimensions of bat biodiversity and the remaining landscape variables after model optimization. Model performance and comparisons are listed in Extended Data Tables 1–3, in which the summarized models are identifiable in the list by bold text. After model selection, best-fit models were rerun and are presented here using restricted maximum likelihood. Additional, less competitive models using island size and isolation distance to predict estimated species richness are also presented for the island ecosystem because of their relevance to island biogeographic theory.

Extended Data Table 5 | Regression coefficients and relevant statistics generated from best-fit models accounting for species richness in the Abundance-Based Assemblage Similarity Index

Abundance-Based Assemblage Similarity Index	Estimate	SE	t-value	p-value
Countryside ecosystem				
(intercept)	0.000	0.020	0.000	1.000
ln(observed number of species)	0.114	0.008	13.810	0.000
Island ecosystem				
(intercept)	0.414	0.035	11.868	0.000
ln(observed number of species)	0.158	0.019	8.381	0.000

The models explain relationships between the Abundance-Based Assemblage Similarity Index and observed species richness for each net (see Methods). Relationships are plotted in Extended Data Fig. 2 and were used to calculate the Assemblage Abundance Shift Index used throughout analyses and results.

Extended Data Table 6 | Observed diurnal roosts in deforested habitats in the countryside ecosystem

Species	Common name	Roost structure	Offspring observed
<i>Glossophaga commissarisi</i> ‡	Commissaris' long-tongued bat	Abandoned building on farm	Yes
<i>Glossophaga soricina</i> *†‡	Common long-tongued bat	Abandoned building on farm	Yes
<i>Carollia castanea</i> ‡	Chestnut short-tailed bat	Abandoned building on farm	Yes
<i>Carollia perspicillata</i> * †‡	Seba's short-tailed bat	Abandoned building on farm	Yes
<i>Carollia sowelli</i> ‡	Sowell's short-tailed bat	Abandoned building on farm	Yes
<i>Artibeus jamaicensis</i> *	Jamaican fruit-eating bat	Live fence of "Caña India" or <i>Dracaena fragrans</i>	No
<i>Artibeus lituratus</i>	Great fruit-eating bat	Ornamental, non-native palm tree on farm	Yes
<i>Uroderma bilobatum</i> *†‡	Common tent-making bat	Live fence of "Caña India" or <i>Dracaena fragrans</i>	Yes
<i>Thyroptera tricolor</i>	Spix's disk-winged bat	<i>Heliconia pendula</i> planted near house	No

* Indicates that roost was found by diurnal location of individuals fitted with radio telemetry transponders.

† Indicates that individuals fitted with radio telemetry transponders were originally captured inside forest fragments, but roosted in deforested habitat.

‡ Indicates that roost was used by the same species at least 12 months after initial discovery.

Casual observations of bat roosts were recorded during bat sampling and related field work in Costa Rica. Some roosts were located using 0.25-g radio telemetry transponders, temporarily attached to some animals in a related study. Radio transponders were temporarily attached using techniques adapted from birds³³. All roosts were found between 24 January and 28 March from 2009 to 2012 and were checked occasionally for use and offspring. The only species whose roost was not used over multiple days was *Thyroptera tricolor*. All roosts hosted multiple individuals of the species listed.

Consequences of biodiversity loss for litter decomposition across biomes

I. Tanya Handa^{1,2}, Rien Aerts³, Frank Berendse⁴, Matty P. Berg³, Andreas Bruder^{5,6}, Olaf Butenschoten⁷, Eric Chauvet^{8,9}, Mark O. Gessner^{5,6,10,11}, Jérémy Jabiol^{8,9}, Marika Makkonen^{3,12}, Brendan G. McKie^{13,14}, Björn Malmqvist[†], Edwin T. H. M. Peeters¹⁵, Stefan Scheu⁷, Bernhard Schmid¹⁶, Jasper van Ruijven⁴, Veronique C. A. Vos⁴ & Stephan Hättenschwiler¹

The decomposition of dead organic matter is a major determinant of carbon and nutrient cycling in ecosystems, and of carbon fluxes between the biosphere and the atmosphere^{1–3}. Decomposition is driven by a vast diversity of organisms that are structured in complex food webs^{2,4}. Identifying the mechanisms underlying the effects of biodiversity on decomposition is critical^{4–6} given the rapid loss of species worldwide and the effects of this loss on human well-being^{7–9}. Yet despite comprehensive syntheses of studies on how biodiversity affects litter decomposition^{4–6,10}, key questions remain, including when, where and how biodiversity has a role and whether general patterns and mechanisms occur across ecosystems and different functional types of organism^{4,9–12}. Here, in field experiments across five terrestrial and aquatic locations, ranging from the subarctic to the tropics, we show that reducing the functional diversity of decomposer organisms and plant litter types slowed the cycling of litter carbon and nitrogen. Moreover, we found evidence of nitrogen transfer from the litter of nitrogen-fixing plants to that of rapidly decomposing plants, but not between other plant functional types, highlighting that specific interactions in litter mixtures control carbon and nitrogen cycling during decomposition. The emergence of this general mechanism and the coherence of patterns across contrasting terrestrial and aquatic ecosystems suggest that biodiversity loss has consistent consequences for litter decomposition and the cycling of major elements on broad spatial scales.

Biological diversity that directly influences litter decomposition exists at multiple trophic levels⁴. This diversity includes plants that produce litter mixtures of varying quality, microbial decomposers and invertebrate consumers of varying body size, the last two of which selectively use the heterogeneous resources provided by litter mixtures^{4,13}. General principles of the effects of biodiversity on litter decomposition have proved elusive: both pioneering work¹⁴ and recent syntheses have highlighted contrasting effects of litter species richness on decomposition^{4–6,15,16}. In part, this variation appears to be due to site-specific conditions, including contrasts between aquatic and terrestrial ecosystems, as well as between geographic settings. Further differences may arise from variation in experimental protocols, the plant species studied and the types of decomposers included in a given experiment. Such methodological discrepancies have complicated syntheses across studies, hindering the emergence of common patterns and mechanisms.

Here we report the results of the first concerted experiments studying the effects of biodiversity on decomposition by manipulating diversity across trophic levels and distinct biomes in both forest floor and

stream habitats (Extended Data Table 1). We proposed that the functional diversity of decomposers (variation in body size) and of leaf litter (variation in litter quality) promotes C and N cycling across contrasting locations (subarctic to tropical) and ecosystem types (terrestrial versus aquatic). Body size encapsulates numerous species traits that are relevant to ecosystem functioning, and extinction scenarios project that the larger species will be preferentially lost from biological communities^{17,18}. Similarly, plant functional types reflect differences in leaf quality traits that determine litter decomposition independently of geographical location¹⁹. Plant functional types are defined here in terms of plant C allocation strategies (deciduous versus evergreen), N acquisition strategies (N-fixing versus non-N-fixing) and litter recalcitrance (rapidly decomposing versus slowly decomposing) (Extended Data Table 2).

Mixing leaf litter from various plant functional types together resulted in accelerated C and N dynamics, as indicated by the overall net positive effects on C and N loss (that is, increased C and N loss with increasing functional diversity) (Fig. 1, C loss; Extended Data Fig. 1, N loss; $P < 0.05$, C and N loss). However, C loss from litter mixtures was only $2.9 \pm 0.8 \text{ mg g}^{-1}$ (mean \pm s.e.m.) of initial litter dry mass greater than the expected loss based on data from single litter functional types, indicating only a modest increase in C cycling as a result of litter mixing. Although also statistically significant ($P < 0.01$), the difference in the loss of N across all litter mixtures was very small ($0.1 \pm 0.2 \text{ mg N g}^{-1}$ of initial litter dry mass; mean \pm s.e.m.; Extended Data Fig. 1). The net litter diversity effect on C loss was stronger in terrestrial than in aquatic ecosystems ($P < 0.001$, Fig. 1 and Extended Data Table 3), supporting theoretical predictions⁴ but contrasting the results of a meta-analysis in which diversity effects on decomposition were significant only for streams⁶. Sorting the litter mixtures into species at the end of the experiments enabled us to explore potential reasons for this discrepancy, by partitioning the net diversity effects into complementarity effects (that is, the effects resulting from synergistic or antagonistic interactions) and selection effects (that is, the effects arising when the presence of a particular functional type with high (or low) process rates dominates a mixture)²⁰. The observed net diversity effects were clearly driven by complementarity effects that were stronger than selection effects (Fig. 1). Overall, the complementarity effect was a similar strength to the net effect for C loss ($3.4 \pm 0.9 \text{ mg C g}^{-1}$), and even stronger than the net effect for N loss ($1.0 \pm 0.2 \text{ mg N g}^{-1}$). By contrast, the mean selection effects were not significant. The characteristics of the forest floor habitat that may favour complementarity effects include strong fluctuations in temperature and humidity and a homogenous litter cover⁴. Conversely,

¹Centre d'Ecologie Fonctionnelle et Evolutive (CEFE), CNRS, 1919 Route de Mende, 34293 Montpellier, France. ²Département des Sciences Biologiques, Université du Québec à Montréal, C.P. 8888, succursale Centre-ville, Montréal, Québec H3C 3P8, Canada. ³Department of Ecological Science, VU University Amsterdam, De Boelelaan 1085, 1081 HV Amsterdam, The Netherlands. ⁴Nature Conservation and Plant Ecology Group, Wageningen University, Droevendaalsesteeg 3a, 6708 PB Wageningen, The Netherlands. ⁵Department of Aquatic Ecology, Eawag: Swiss Federal Institute of Aquatic Science and Technology, Überlandstrasse 133, 8600 Dübendorf, Switzerland. ⁶Institute of Integrative Biology (IBZ), ETH Zürich, 8092 Zürich, Switzerland. ⁷Georg August University Göttingen, J.F. Blumenbach Institute of Zoology and Anthropology, Berliner Strasse 28, 37073 Göttingen, Germany. ⁸Université de Toulouse, INP, UPS, EcoLab (Laboratoire Ecologie Fonctionnelle et Environnement), 118 Route de Narbonne, 31062 Toulouse Cedex, France. ⁹CNRS, EcoLab, 118 Route de Narbonne, 31062 Toulouse Cedex, France. ¹⁰Leibniz Institute of Freshwater Ecology and Inland Fisheries (IGB), Alte Fischerhütte 2, 16775 Stechlin, Germany. ¹¹Department of Ecology, Berlin Institute of Technology (TU Berlin), Ernst-Reuter-Platz 1, 10587 Berlin, Germany. ¹²Climate Change Programme, Finnish Environment Institute, PO Box 140, 00251 Helsinki, Finland. ¹³Department of Ecology and Environmental Science, Umeå University, 90187 Umeå, Sweden. ¹⁴Department of Aquatic Sciences and Assessment, Swedish University of Agricultural Sciences, PO Box 7050, 75007 Uppsala, Sweden. ¹⁵Aquatic Ecology and Water Quality Management Group, Wageningen University, PO Box 47, 6700 AA Wageningen, The Netherlands. ¹⁶Institute of Evolutionary Biology and Environmental Studies & Zürich-Basel Plant Science Center, University of Zürich, Winterthurerstrasse 190, 8057 Zürich, Switzerland. [†]Deceased.

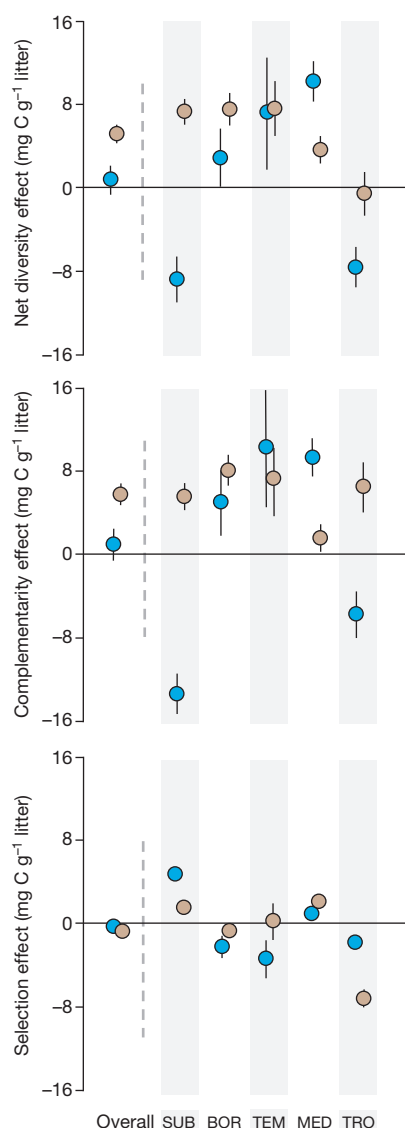


Figure 1 | Net diversity, complementarity and selection effects of plant litter mixtures on C loss. The net diversity effect is the deviation from the expected mean based on C loss measured from litter consisting of single species. The blue and brown circles show the mean effects (\pm s.e.m.) on C loss from litter mixtures in forest streams and on forest floors, respectively, in subarctic (SUB), boreal (BOR), temperate (TEM), Mediterranean (MED) and tropical (TRO) locations. Each circle to the right of the dashed lines shows the mean effect per ecosystem type (that is, aquatic versus terrestrial), as calculated across the three types of decomposer community ($n = 165$ litter mixtures per location and ecosystem type; see Extended Data Table 3 for statistical analyses). The circles to the left of the dashed lines show the overall mean across all locations.

the observed negative complementarity effects in subarctic and tropical streams could reflect a low density and low taxon richness of litter consumers (Extended Data Table 4) and thus limited potential for complementary resource use²¹.

Our experiments also show that completeness of the decomposer community, which is rarely considered in large-scale studies, is important for C and N dynamics during litter decomposition (Fig. 2, Table 1 and Extended Data Table 5). The presence of medium-sized invertebrates (≤ 1 mm in diameter) in the decomposer community increased the average C and N loss across all sites by $2.1 \pm 0.8\%$ and $2.0 \pm 1.0\%$, respectively. The complete decomposer community (which included organisms up to 5 mm in diameter) increased the average C loss across all sites by $10.6 \pm 1.0\%$ and the average N loss across all sites by $11.1 \pm 1.2\%$ (Fig. 2). This effect was consistently positive across all but the Mediterranean

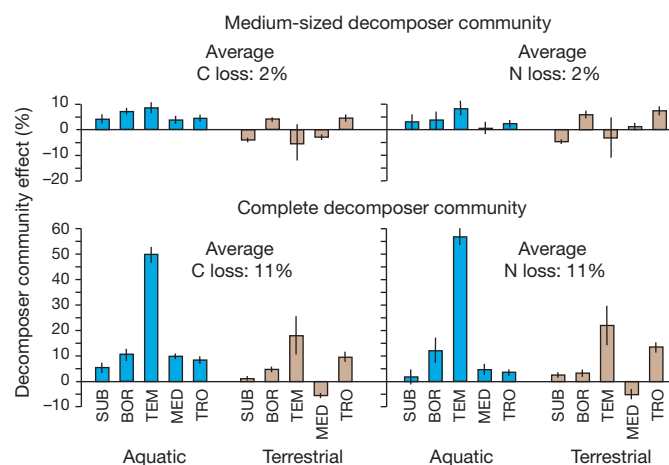


Figure 2 | Effect of decomposer community completeness on litter C and N loss. C loss (left) and N loss (right) from all litter treatments (all single species and all mixtures) exposed to medium-sized decomposers (top; percentage difference compared with the smallest mesh size) and the complete decomposer community (bottom; percentage difference compared with the smallest mesh size). The blue and brown bars show mean effects (\pm s.e.m.) in forest streams and on forest floors, respectively, in the five indicated locations ($n = 45$ litter treatments per location per ecosystem type; see Table 1 for statistical analyses).

terrestrial site. Thus, the presence of large fauna clearly has a major impact on decomposition (Table 1), as reported previously^{22–24}; however, in line with previous studies, the importance of large fauna varies among locations in aquatic²² and terrestrial ecosystems^{23,24}. In our study, the strong effects of the complete decomposer community at the temperate and tropical locations correspond to high relative abundances of millipedes and termites at the terrestrial temperate and tropical sites, respectively (Extended Data Table 6). Similarly, the large effect of the complete decomposer community at the temperate aquatic site corresponds to the high abundance of a particularly efficient amphipod detritivore (Extended Data Table 4). Our data clearly indicate that the large-bodied organisms are the most critical for decomposition. These animals also tend to face the greatest extinction risk¹⁷.

Litter mixing and completeness of the decomposer community interacted with each other to affect C and N loss, although this interaction explains less of the variance than the main effects (Table 1). C loss and, even more so, N loss increased in the presence of particular plant functional types and with increasing completeness of the decomposer community (Table 1 and Extended Data Table 7). Although the type of decomposer community did not significantly change the net effect of diversity on C loss ($P = 0.67$) or N loss ($P = 0.30$) (Extended Data Table 3), it emerged as a significant factor in the selection effect for both C loss ($P < 0.05$) and N loss ($P < 0.05$). Additionally, the interaction between the rapidly decomposing litter type and the decomposer community was significant in explaining the selection effect and the overall net diversity effect on C loss ($P < 0.05$) and N loss ($P < 0.05$), suggesting that large decomposers are particularly important drivers of C and N loss from litter mixtures that contain rapidly decomposing litter. The food preference behaviour of decomposers could be important in accounting for this result, as has previously been implied for terrestrial²⁵ and aquatic²⁶ ecosystems.

A key result of our large-scale study is that the effects of litter diversity on C and N dynamics can be largely explained by the presence of particular functional plant types in litter mixtures, supporting the idea that the range and relative abundance of plant traits in ecosystems underlie the effects of species richness on ecosystem processes^{27,28}. The effects of the presence of litter from particular plant functional types, or the interactions among these, were consistent across locations at both terrestrial and aquatic sites, together accounting for about 10% of the total variance as shown in the full analysis of variance model (Table 1, $P < 0.05$;

Table 1 | Variance in C and N loss associated with diversity and sites

Factor	DF	C loss			N loss		
		%SS	F	P	%SS	F	P
Variation associated with diversity							
Litter community: litter richness*	1	0	0	0.93	1.0	63.7	***
Litter community: remaining variation associated with FT composition†	13	10.1	66.0	***	8.9	44.4	***
Decomposer community (small, medium-sized or complete)	2	5.8	247	***	4.4	142	***
Litter community × decomposer community	28	0.8	2.45	***	1.3	3.1	***
Variation associated with sites							
Location (tropical, Mediterranean, temperate, boreal or subarctic)	4	12.1	52.8	***	16.9	66.5	***
Ecosystem (aquatic stream or terrestrial forest floor)	1	6.4	113	***	1.4	22.3	***
Location × ecosystem	4	6.3	27.7	***	8.0	31.6	***
Block (within location × ecosystem)	40	2.3	4.87	***	2.5	4.13	***
Variation associated with diversity or sites							
Decomposer community × location	8	8.6	92.1	***	7.5	61.3	***
Remaining variance	388	27.1	5.95	***	21.4	3.60	***
Total variance explained by the model	489	79.5	13.9	***	73.3	9.76	***
Residuals	1,739	20.4	—	—	26.7	—	—
Total	2,228	100.0	—	—	100.0	—	—

The relative contributions of variance in C and N loss associated with diversity and sites (expressed in percentage sums of squares (% SS)) in a large-scale leaf litter decomposition experiment. The main factors are italicized. ***, $P < 0.001$, analysis of variance based on sequential sums of squares (see Methods). See Extended Data Table 5 for details. DF, degrees of freedom; FT, functional type.

* Plant species (trees or shrubs) were selected to represent the same four functional types at each location (N-fixing, evergreen, rapidly decomposing deciduous and slowly decomposing deciduous). Linear functional type richness was fitted before litter functional type compositions.

† An alternative model omitting richness and testing in detail the litter functional type compositions in a full factorial analysis with contrasts for functional type presence/absence and interactions is presented in Extended Data Table 7. That model highlights the importance of the interaction between the litter of the N-fixing functional type and the rapidly decomposing functional type, hinting at a N-transfer mechanism.

Extended Data Table 7). Beyond the presence or absence of particular plant functional types, we found no significant effect of the richness of plant functional types in the litter on C loss ($P = 0.93$), although a positive effect was observed on N loss ($P < 0.001$) (Table 1). The effect on N loss was strongest when the most complete decomposer communities had access to the litter (litter richness × decomposer community interaction; $P < 0.05$). Our results indicate that partitioning the diversity effects into the separate contributions of the presence or absence of particular plant functional types in litter and their interactions can help move interpretations of biodiversity–ecosystem functioning experiments beyond the current dichotomy between broad generalizations and claims of idiosyncratic compositional effects^{5,14,15}.

An intriguing finding in this context is that the strongest positive interaction emerged between two particular litter functional types: N-fixing plants and rapidly decomposing deciduous plants (Extended Data Table 7). When these types were present together in litter mixtures, the average C loss was 13.5% greater than the average C loss of all litter combinations, and the N loss was 32.5% greater. This general pattern holds across

terrestrial and aquatic ecosystems from the subarctic to the tropics. Moreover, relative to the total amount of N in the litter initially, less N remained in the litter of N-fixing plants when rapidly decomposing litter was present than when it decomposed alone (Fig. 3 and Extended Data Table 8). The rapidly decomposing litter, in turn, contained more N when litter from N-fixing plants was present than when it decomposed alone (Fig. 3 and Extended Data Table 8). On average across all of the sites, the litter of N-fixing plants lost 20.6% of its initial N when it decomposed alone but 25.0% when it decomposed in the presence of litter from rapidly decomposing plants. By contrast, the litter of rapidly decomposing plants lost 18.1% of its N when it decomposed alone but 13.4% when litter from N-fixing plants was present. This striking pattern across locations and ecosystems suggests, for the first time from field data, that N can be transferred between litter types. A plausible mechanism for this effect is that fungal decomposers tap the nutrient reservoir of the N-fixing plant litter, boosting C use and fungal growth in the N-deficient litter, which provide high-quality C²⁹ (see section Extended discussion on litter N transfer in Methods). The average net differences in N fluxes between single-species litter and litter mixtures of these two plant functional types account for approximately 0.25 g N per square metre of ground area, representing up to one-tenth of the total annual N input from leaf litter fall. Thus, although the biodiversity effects that we report here, in line with recent syntheses^{9–11}, are smaller than those noted for other ecosystem processes such as plant biomass production, these changes in N fluxes can have important consequences for the ecosystem. Even slight differences in the N dynamics in litter mixtures compared with the respective single-species litter can substantially change the N supply to primary producers and other organisms over large spatial and temporal scales³⁰.

The implications of our results are that changes in C and N cycling in response to biodiversity loss are largely predictable across vastly different latitudes in both terrestrial and aquatic ecosystems, by taking into account relatively simple plant traits and the structural characteristics of decomposer communities. To provide robust projections of how ecosystems respond to a loss of biodiversity, it is essential to identify the mechanisms that result from specific interactions between the components of biodiversity as we describe here. With the consistent patterns and mechanisms of biodiversity effects that we have shown, such projections now seem to be within reach.

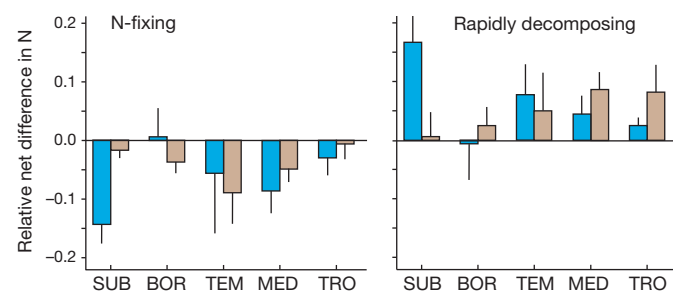


Figure 3 | Relative change in the total amount of litter N. The relative net difference between two-species mixtures (containing litter from the N-fixing and the rapidly decomposing plant) and monocultures of N-fixing plant litter (left) or rapidly decomposing plant litter (right) is shown (mean ± s.e.m., $n = 15$; see Extended Data Table 8 for statistical analyses) for litter decomposing in terrestrial (brown) and aquatic (blue) ecosystems at five locations. The relative net difference was calculated as $[(N_{i,m} - N_{f,m})/N_{i,m}] - [(N_{i,a} - N_{f,a})/N_{i,a}]$, where $N_{i,m}$ and $N_{i,a}$ are the initial (i), and $N_{f,m}$ and $N_{f,a}$ are the final (f), amounts of N in a particular litter type in a mixture (m) or alone (a).

METHODS SUMMARY

The field experiments followed an identical protocol at ten sites, encompassing both aquatic (forest stream) and terrestrial (forest floor) ecosystems at five locations across a latitudinal gradient spanning from the subarctic to the tropics, with intermediate locations in boreal, temperate and Mediterranean climates (Extended Data Table 1). Leaf litter from native tree or shrub species representing four common functional types (evergreen, deciduous with slowly decomposing litter, deciduous with rapidly decomposing litter, and N-fixing) that naturally occur across all locations (18 species in total; Extended Data Table 2) was exposed to decomposers in a total of 2,250 experimental microcosms set up in the field with all possible location-specific single-species and multi-species combinations. We used a randomized block design with five blocks per site. Each block contained 1 replicate of 15 combinations of litter types (that is, all possible combinations of 4 litter species) \times 3 microcosm mesh sizes (totalling 45 microcosms per block). The three mesh sizes used to construct the field microcosms allowed us to establish three increasingly complete decomposer communities (small, medium-sized and complete) in the microcosms. Small-sized decomposer communities included microorganisms and fauna that passed through 50- μ m and 250- μ m mesh screens (DIATEX) in terrestrial and aquatic ecosystems, respectively. The medium-sized decomposer communities contained all organisms (including invertebrates) that passed through 1-mm mesh screens, whereas the complete decomposer communities included all organisms that passed through 5-mm mesh screens. Litter mass loss was allowed to proceed to the same defined decomposition stage (40–50% of the mass of the least recalcitrant litter type remaining at each location; Extended Data Table 9) to ensure the comparisons of C and N loss, as well as the effects of diversity, at a similar decomposition stage among sites, using analysis of variance models (see Methods section).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 28 November 2013; accepted 11 March 2014.

- Wardle, D. A. *Communities and Ecosystems: Linking the Aboveground and Belowground Components* (Princeton Univ. Press, 2002).
- Bardgett, R. D. *The Biology of Soil: A Community and Ecosystem Approach* (Oxford Univ. Press, 2005).
- Parton, W. *et al.* Global-scale similarities in nitrogen release patterns during long-term decomposition. *Science* **315**, 361–364 (2007).
- Gessner, M. O. *et al.* Diversity meets decomposition. *Trends Ecol. Evol.* **25**, 372–380 (2010).
- Hättenschwiler, S., Tiunov, A. V. & Scheu, S. Biodiversity and litter decomposition in terrestrial ecosystems. *Annu. Rev. Ecol. Syst.* **36**, 191–218 (2005).
- Cardinale, B. J. *et al.* The functional role of producer diversity in ecosystems. *Am. J. Bot.* **98**, 572–592 (2011).
- May, R. M. Why should we be concerned about loss of biodiversity. *C. R. Biol.* **334**, 346–350 (2011).
- Naeem, S., Duffy, J. E. & Zavaleta, E. The functions of biological diversity in an age of extinction. *Science* **336**, 1401–1406 (2012).
- Cardinale, B. J. *et al.* Biodiversity loss and its impact on humanity. *Nature* **486**, 59–67 (2012).
- Hooper, D. U. *et al.* A global synthesis reveals biodiversity loss as a major driver of ecosystem change. *Nature* **486**, 105–108 (2012).
- Balvanera, P. *et al.* Quantifying the evidence for biodiversity effects on ecosystem functioning and services. *Ecol. Lett.* **9**, 1146–1156 (2006).
- Loreau, M. Linking biodiversity and ecosystems: towards a unifying ecological theory. *Phil. Trans. R. Soc. B* **365**, 49–60 (2010).
- Reiss, J., Bridle, J. R., Montoya, J. M. & Woodward, G. Emerging horizons in biodiversity research and ecosystem functioning. *Trends Ecol. Evol.* **24**, 505–514 (2009).
- Wardle, D. A., Bonner, K. I. & Nicholson, K. S. Biodiversity and plant litter: experimental evidence which does not support the view that enhanced species richness improves ecosystem function. *Oikos* **79**, 247–258 (1997).
- Gartner, T. B. & Cardon, Z. G. Decomposition dynamics in mixed-species leaf litter. *Oikos* **104**, 230–246 (2004).
- Lecerf, A. *et al.* Incubation time, functional litter diversity, and habitat characteristics predict litter-mixing effects on decomposition. *Ecology* **92**, 160–169 (2011).
- Duffy, J. E. Biodiversity loss, trophic skew and ecosystem functioning. *Ecol. Lett.* **6**, 680–687 (2003).
- Woodward, G. *et al.* Body size in ecological networks. *Trends Ecol. Evol.* **20**, 402–409 (2005).
- Cornwell, W. K. *et al.* Plant species traits are the predominant control on litter decomposition rates within biomes worldwide. *Ecol. Lett.* **11**, 1065–1071 (2008).
- Loreau, M. & Hector, A. Partitioning selection and complementarity in biodiversity experiments. *Nature* **412**, 72–76 (2001).
- Frainer, A., McKie, B. G. & Malmqvist, B. When does diversity matter? Species functional diversity and ecosystem functioning across habitats and seasons in a field experiment. *J. Anim. Ecol.* **83**, 460–469 (2014).
- Woodward, G. *et al.* Continental-scale effects of nutrient pollution on stream ecosystem functioning. *Science* **336**, 1438–1440 (2012).
- Wall, D. H. *et al.* Global decomposition experiment shows soil animal impacts on decomposition are climate-dependent. *Glob. Chang. Biol.* **14**, 2661–2677 (2008).
- García-Palacios, P., Maestre, F. T., Kattge, J. & Wall, D. H. Climate and litter quality differently modulate the effects of soil fauna on litter decomposition across biomes. *Ecol. Lett.* **16**, 1045–1053 (2013).
- Vos, V. C. A., van Ruijven, J., Berg, M. P., Peeters, E. T. H. M. & Berendse, F. Macro-detritivore identity drives leaf litter diversity effects. *Oikos* **120**, 1092–1098 (2011).
- Swan, C. M. & Palmer, M. A. Preferential feeding by an aquatic detritivore mediates non-additive decomposition of speciose leaf litter. *Oecologia* **149**, 107–114 (2006).
- Garnier, E. *et al.* Plant functional markers capture ecosystem properties during secondary succession. *Ecology* **85**, 2630–2637 (2004).
- Cadotte, M. W., Carscadden, K. & Mirotchnick, N. Beyond species: functional diversity and the maintenance of ecological processes and services. *J. Appl. Ecol.* **48**, 1079–1087 (2011).
- Schimel, J. P. & Hättenschwiler, S. Nitrogen transfer between decomposing leaves of different N status. *Soil Biol. Biochem.* **39**, 1428–1436 (2007).
- Finzi, A. C. & Canham, C. D. Non-additive effects of litter mixtures on net N mineralization in a southern New England forest. *For. Ecol. Manage.* **105**, 129–136 (1998).

Acknowledgements We thank A. Lecerf and P. García-Palacios for comments on the manuscript. We are grateful to numerous technicians in Montpellier (France) for building field microcosms, in Dübendorf (Switzerland) for water chemical analyses, and in Dübendorf and Göttingen (Germany) for grinding litter samples. We also thank M. Schindler for assistance, B. Buatois, R. Leclerc, P. Schevin and L. Sonié for analyses performed at the Plate-Forme d'Analyses Chimiques en Ecologie, LabEx CeMEB (France), G. Larocque for help with R code and our many colleagues at the field sites and research institutes for their support in various ways. This study is part of the BioCycle research project funded by the European Science Foundation (ESF) as part of its EUROCORES programme EuroDIVERSITY. BioCycle has been endorsed by DIVERSITAS as contributing towards their scientific research priorities in biodiversity science.

Author Contributions All authors contributed to the experimental design, data acquisition and revision of the final manuscript. Statistical analyses were performed by I.T.H., B.S., J.V.R. and B.G.M., and the manuscript was written by I.T.H., S.H. and M.O.G.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to S.H. (stephan.hattenschwiler@cefe.cnrs.fr).

METHODS

Experimental design. Our field experiments followed an identical protocol at a total of ten sites, representing either an aquatic ecosystem (forest stream) or a terrestrial ecosystem (forest floor). Five locations were selected across a broad latitudinal gradient spanning from the subarctic to the tropics, with intermediate locations in boreal, temperate and Mediterranean climates (Extended Data Table 1). Across all five locations, and in both the stream and forest ecosystems, the experiments consisted of a randomized block design in which the leaf litter from 4 common native plants (corresponding to the functional types shown in Extended Data Table 2) and 11 mixtures of these litter types (corresponding to all possible litter combinations within a location) were enclosed in nylon mesh screens (DIATEX) and placed in the field in five blocks ($n = 5$ locations $\times 2$ ecosystem types $\times 15$ litter combinations $\times 3$ mesh sizes $\times 5$ blocks = 2,250 microcosms). The four functional plant types represent distinct plant C allocation strategies (deciduous versus evergreen), N acquisition strategies (N fixer versus not a N fixer) and litter recalcitrance of deciduous non-N-fixers (rapidly decomposing versus slowly decomposing).

The three mesh sizes used to construct the microcosms enabled us to distinguish three increasingly complete decomposer communities (small, medium-sized and complete) that established themselves on the decomposing litter. Small decomposers included microorganisms and small-sized fauna that passed through 50- μm and 250- μm mesh screens in terrestrial and aquatic systems, respectively. The medium-sized decomposer communities contained all organisms (including invertebrates) that passed through 1-mm mesh screens, whereas the complete decomposer communities included all decomposers that passed through 5-mm mesh screens. Litter mass loss was allowed to proceed to the same defined decomposition stage (40–50% of the litter mass of the least recalcitrant litter type remaining at each site; Extended Data Table 9) to ensure meaningful comparisons of C and N loss among all sites. At all ten sites, extra microcosms containing the fastest decomposing litter type served as benchmark indicators of decomposition rates.

Site characterization. The five stream locations were characterized in terms of their geomorphological, physical and chemical features (Extended Data Table 1). Water samples were collected for chemical analyses at the time of establishing the experiments. Samples for inorganic N and phosphorus determination were filtered over 0.45- μm pore-size cellulose acetate membrane filters and transported to the laboratory in a cooler at about 5 °C, where they were frozen for later analysis at Eawag, Switzerland.

At the five forest sites (Extended Data Table 1), the leaf area index was measured at breast height on a uniformly cloud-covered day when the forest canopy was fully developed, using an LAI-2000 Plant Canopy Analyzer (LI-COR) for the subarctic location, an LAI-2200 Plant Canopy Analyzer (LI-COR) for the Mediterranean and tropical locations, and a SunScan Canopy Analysis System (Delta T Devices) for the temperate locations. Data for the boreal location were provided by K. Bishop & P.-E. Mellander. To characterize the soil at each of the terrestrial sites, three samples from each experimental block were taken with a soil corer (5 cm diameter, 10 cm height), pooled, stored in plastic bags at 4 °C, then sent cooled to the University of Göttingen (Germany). Sieved soil samples (<2 mm sieve pore size) were analysed for pH (2 g soil in 20 ml 0.01 M CaCl_2) and C and N concentration (using an NA 1500 Carlo Erba elemental analyser). The microbial biomass of the soil was estimated using the substrate-induced respiration (SIR) method. The microbial respiratory response was measured in an electrolytic O_2 -microcompensation apparatus at 22 °C. These measurements were made hourly for 24 h. The microbial biomass was measured after the addition of glucose (8 mg C g^{-1} dry soil) as a substrate to saturate the catabolic activity of the microorganisms. The maximum initial respiratory response (MIRR; $\mu\text{l O}_2 \text{ g}^{-1}$ dry mass h^{-1}) was calculated as the mean of the lowest three readings within the first 10 h, and the microbial biomass was calculated as $C_{\text{mic}} = 38 \times \text{MIRR}$ (in $\mu\text{g C}_{\text{mic}} \text{ g}^{-1}$ soil dry mass).

Data loggers (SL52T, Signatrol) were installed in some microcosms at all ten sites, to record the temperature every 2 h. These temperature measurements were taken in the same litter treatment for all three mesh sizes in three of the five experimental blocks.

Leaf litter collection. A total of 20 litter types were collected at the 5 locations of our coordinated experiment. This litter corresponded to the same four functional types per location that were introduced above: N-fixing plants, rapidly decomposing deciduous plants, slowly decomposing deciduous plants and evergreen plants (Extended Data Table 2). Litter from these four functional types varies with respect to several quality traits³¹ (Extended Data Table 2). The selected species were common native trees or, in two cases, native woody shrubs (*Vaccinium vitis-idaea* and *Rhododendron tomentosum*) occurring at each location. The litter was collected during location-specific leaf senescence either by hand (*V. vitis-idaea* and *R. tomentosum*) or by using litter traps. An exception was litter from the temperate evergreen species *Ilex aquifolium*, which was obtained by cutting branches in the field and simulating senescence in the laboratory for three weeks. Leaves with signs of herbivory

or disease were discarded. The litter from multiple individual trees or shrubs of each species was pooled and dried at 40 °C.

Leaf litter field incubations. Stream experiments were conducted by exposing 5 g litter batches in tetrahedral mesh microcosms (17 cm \times 25 cm) made of one of three mesh sizes (250 μm , 1 mm or 5 mm) to provide access to decomposer communities differing in body size. The microcosms were randomly attached (about 40 cm distance between microcosms) to five 20-m metal chains, each in a separate riffle (short, relatively shallow sections of streams with non-stagnating water) 20 m or farther apart from each other (experimental blocks). The chains were fixed in the stream with reinforcing bars in fairly homogeneous sand-gravel stream sections, where leaves accumulated naturally. All microcosms were submerged at depths sufficient to ensure that they were not exposed to air when water levels dropped. Care was taken to expose the litter to constant flow conditions, avoiding deep depositional areas (that is, pools and backwaters) with slow or no flow and rocky riffles with broken flow.

Terrestrial experiments on the forest floor were conducted by incubating 8 g location-specific litter (4 g only in the subarctic because of limited litter availability for some species) in field microcosms made of polyethylene cylinders (15 cm diameter, 10 cm height) covered with 50- μm mesh at the top and bottom to allow the passage of water but to prevent the entry of natural litter fall from above and the loss of small litter particles from the bottom. Two windows (5 cm \times 18 cm) were cut into the cylinders and covered with 50- μm , 1-mm or 5-mm mesh to provide access to decomposer communities differing in body size. The windows were cut close to the bottom of the cylinders to ensure that decomposers had access to a continuous layer of litter outside and inside the microcosms. An additional 1.5-cm height plastic ring of the same diameter as the cylindrical microcosms was attached at the bottom of the microcosm, making it possible to push the microcosms gently into the top soil (to a depth of 1.5 cm). This ring held the terrestrial microcosms properly in place while the bottom mesh was in intimate contact with the soil surface. In cases in which pushing the microcosms into the soil was difficult (for example, in the tropical forest with its dense superficial tree roots), the 1.5-cm rings were fitted with a separate plastic or metal ring before placing the microcosms. The microcosms were separated from each other by at least 50 cm. They were randomly distributed within blocks that were established at least 20 m apart from each other.

Sample harvest and processing. We removed the decomposing litter of all species from the field when 40–50% of the initial litter mass of the fastest decomposing species was remaining. As a consequence, the duration of litter decomposition varied among locations and ecosystem types (Extended Data Table 9). This procedure ensured that similar decomposition stages were sampled at all sites, facilitating meaningful comparisons of decomposition rates and litter diversity effects. All litter samples were separated into the constituent species immediately after litter retrieval. The litter recovered from the streams was gently washed to remove any adhering material and invertebrates. The litter from the terrestrial sites was cleaned by gently brushing off any dirt without using water, to prevent the leaching of nutrients. The litter samples were then dried at 65 °C for 48 h. A correction factor was used to convert the initial litter mass (weighed after drying at 40 °C) to the final dry mass, based on ten randomly selected samples per litter type that were successively dried and weighed in the laboratory first at 40 °C and then at 65 °C.

Litter C and N loss. The initial C and N concentrations of each of the 20 individual litter types were determined from 5 random samples. The final C and N concentrations after retrieval of the litter from the field were also measured for each individually sorted litter type from each microcosm. This process resulted in a total of 5,400 samples for which to calculate the percentage C and N loss for each litter type under the various conditions. Following the determination of litter dry mass, all initial and final samples were ground with a ball mill (Retsch PM 400) to a fine homogeneous powder. Subsamples of 3 mg were analysed for C and N concentrations using a CHN elemental analyser (Flash EA 1112 Series, Thermo Finnigan). C and N loss (%) from the litter during field exposure was calculated as $100 \times [(M_i \times CN_f) - (M_f \times CN_i)] / (M_i \times CN_i)$, where M_i and M_f are the initial and final litter dry mass, respectively, and CN_i and CN_f are the initial and final C or N concentration (% of litter dry mass). Using C loss (%) rather than total litter mass loss allowed us to correct for any possible inorganic contamination of the litter retrieved from the field.

Analyses of diversity effects and statistical models. The net diversity effects, comprising complementarity and selection effects, on both C and N loss were calculated in species mixtures²⁰. The net diversity effect was calculated as the sum of the complementarity and selection effects and contrasts the actual C and N loss observed for mixtures of plant functional types with that expected based on the C and N loss measured in single-species treatments. The net diversity effect represents the sum of synergistic or antagonistic interactions (that is, complementarity effects) and those due to the presence of a dominant species (that is, selection effects). Data were square-root transformed (keeping the original negative and positive signs

for the transformed values) to meet the assumptions for the analysis of variance of net diversity, complementarity and selection effects (see details below).

Analysis of variance models based on sequential sums of squares (type I) were used to assess the effects of diversity (the richness of plant litter functional types or the presence or absence of a given functional type and its interaction with other functional types), the completeness of the decomposer community (small, medium and large (complete)), the location across the latitudinal gradient and the ecosystem type (terrestrial versus aquatic) on percentage C and N loss. To ensure meaningful comparisons across the locations, several standardization methods were tested to remove any variation associated with the differences in incubation length. These methods included standardizing relative to the following: 1) a standard litter type from a non-native plant, *Ailanthus altissima*, that decomposed at all locations during the experiments; 2) the overall mean C or N loss per mesh size across locations; and 3) the mean C or N loss per mesh size of the rapidly decomposing functional type across locations. Because the results were consistent irrespective of standardization, the final model is presented using the non-standardized data.

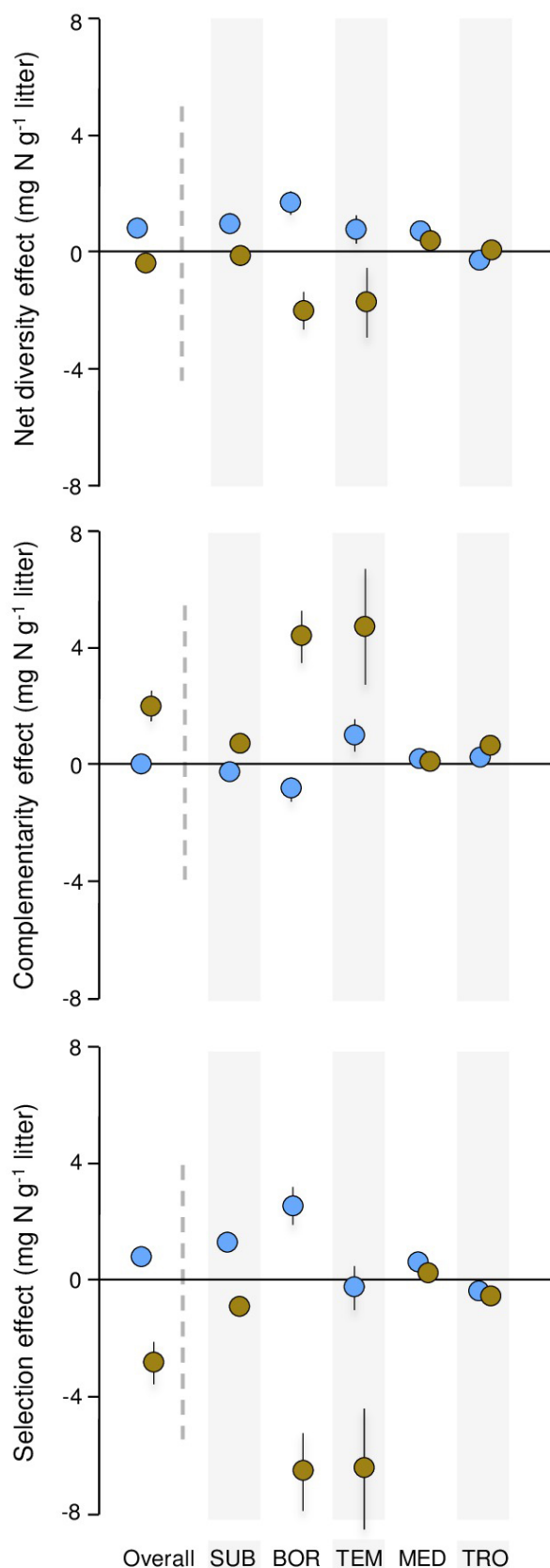
The model terms were fitted to account for the dependency between the richness of plant litter functional types and the functional type composition (the presence or absence of a given functional type and interactions between functional types). First, functional type composition was partitioned into a contrast for richness and residual functional type composition (Table 1 and Extended Data Table 5). Second, as shown in Extended Data Table 7, we omitted the richness term and instead resolved the functional type composition into a full factorial analysis with contrasts for functional type presence or absence and interactions. In this model, the decomposer community was fitted as a log-linear contrast (small to large mesh size was expressed as the logarithm of the mesh size of the microcosms, which produced a linear relationship of the three mesh sizes). We also removed all of the other non-significant interaction terms in multiple successive model-fitting steps. These two alternative analyses reflect different partitionings of the functional type composition term into contrasts; they allowed us to compare the explanatory power of the richness contrast with the presence/absence contrast. A perfect linear richness effect would be found if all presence/absence contrasts had equal coefficients and did not interact. In this case, the mean squares or the richness effect with only one degree of freedom would be much larger than that of the combined mean squares of the presence/absence main effects of the four litter types with four degrees of freedom. In both models, the terms 'location' and 'ecosystem type' were tested at the block level. All other terms were tested against the residuals.

A similar analysis of variance approach was used to test independently for the effects of these same factors on complementarity and selection effects, as well as on net diversity effects (Extended Data Table 3). In a separate analysis of variance (Extended Data Table 8), we also tested whether the net loss of the total amount of N relative to the initial amount of N differed when litter of particular plant functional types (for example, rapidly decomposing litter and litter of N-fixing plants) decomposed together as opposed to decomposing separately, which we interpreted as an indication of N transfer between litter species. The location and ecosystem type were also included in this analysis. All statistical analyses were performed with R software version 2.8.0.

Extended discussion on litter N transfer. Although our data suggest that N was transferred from the litter of N-fixing plants to rapidly decomposing litter, alternative mechanisms cannot be entirely ruled out. In particular, N incorporated into decomposing litter can originate not only from another co-occurring litter type but also from the N pool in the soil or stream water or from microbial N fixation³². However, N transfer from such alternative N sources does not readily explain the concomitant reciprocal changes that we observed between the litter of N-fixing plants and rapidly decomposing litter. Moreover, the idea that N transfer occurred between the two litter types is further supported by a positive net diversity effect on C loss that we observed only when these two particular litter functional types were both present (Extended Data Table 3). Additional support for our interpretation comes from ¹⁵N tracer studies in microcosms with tropical²⁸ and temperate³³ forest litter, which are proof of principle that active biological transfer of N through microorganisms, particularly saprotrophic fungi, can occur. Our large-scale field experiment suggests that this phenomenon might be widespread across terrestrial and aquatic ecosystems and across a wide variety of forest types and climatic conditions.

It had been proposed that N transfer is driven by a gradient in N concentration between litter types^{4,5}, the rationale being that the element that limits the decomposition rate is N. However, the scenario now unfolding from our experiment (Fig. 3) and the recent isotope tracer studies under laboratory conditions^{28,33} is that N transfer is stoichiometrically controlled. The crucial determinant that defines the gradient along which N will be transferred in litter mixtures seems to be the demand for N relative to the availability of C (and possibly that of other elements critical for decomposer growth), rather than differences in the N concentration. A litter with high C quality favours rapid microbial growth, which in turn entails a high demand for N (and other nutrients), resulting in N acquisition from neighbouring nutrient pools. In extreme cases, the N source litter may even have a lower N concentration than the N sink litter³³, provided that the C quality of both litter types is sufficiently different. In accordance with this mechanism, the decomposition of recalcitrant litter types in our study (slowly decomposing and evergreen plant functional types) was not accelerated by the presence of litter from N-fixing plant species (Extended Data Table 7), although those recalcitrant litter types had similarly low or lower initial N concentrations than the rapidly decomposing litter species, which consistently benefited from the presence of N-fixing plant litter.

31. Makkonen, M. *et al.* Highly consistent effects of plant litter identity and functional traits on decomposition across a latitudinal gradient. *Ecol. Lett.* **15**, 1033–1041 (2012).
32. Vitousek, P. M. & Hobbie, S. Heterotrophic nitrogen fixation in decomposing litter: patterns and regulation. *Ecology* **81**, 2366–2376 (2000).
33. Lummer, D., Scheu, S. & Butenschön, O. Connecting litter quality, microbial community and nitrogen transfer mechanisms in decomposing litter mixtures. *Oikos* **121**, 1649–1655 (2012).
34. Lepori, F. & Malmqvist, B. Deterministic control on community assembly peaks at intermediate levels of disturbance. *Oikos* **118**, 471–479 (2009).
35. Nijboer, R. *De Springendalse Beek. Macrofaunagemeenschappen in de Periode 1970–1995* IBN-rapport 455 (Instituut voor Bos- en Natuuronderzoek, Wageningen, 1999).



Extended Data Figure 1 | Net diversity, complementarity and selection effects of plant litter mixtures on N loss. The net diversity effect is the deviation from the expected mean based on N loss measured from litter consisting of single species. The blue and brown circles show the mean effects (\pm s.e.m.) on N loss from litter mixtures in forest streams and on forest floors, respectively, in subarctic (SUB), boreal (BOR), temperate (TEM), Mediterranean (MED) and tropical (TRO) locations. Each circle to the right of the dashed lines shows the mean effect per ecosystem type (that is, aquatic versus terrestrial), as calculated across the three types of decomposer communities ($n = 165$ litter mixtures per location and ecosystem type; see Extended Data Table 3 for statistical analyses). The circles to the left of the dashed lines show the overall mean across all locations ($n = 825$ litter mixtures per ecosystem type).

Extended Data Table 1 | Characteristics of aquatic and terrestrial ecosystems at five widely dispersed locations

Aquatic Locations	Subarctic	Boreal	Temperate	Mediterranean	Tropical
Site	Kopperåsen Sweden	Krycklan Sweden	Mosbeek The Netherlands	Maureillas France	Petit Saut French Guiana
Coordinates	68°26' N, 18°28' E	64°16' N, 19°50' E	52°26' N, 5°32' E	42°28' N, 2°48' E	5°04' N, 53°00' W
Dominant riparian tree cover	<i>Salix</i> spp., <i>Sorbus aucuparia</i>	<i>Picea abies</i> , <i>Betula pubescens</i> <i>Alnus incana</i>	<i>Fagus sylvatica</i>	<i>Quercus ilex</i> , <i>Corylus avellana</i> , <i>Alnus glutinosa</i>	~140 different canopy species
Altitude (m a.s.l.)	445	200	37	180	35
Catchment (km ²)	1.1	19.2	4.9	18.5	0.6
Distance to source (km)	1.6	4.0	1.3	8.8	1.0
Stream order	1	3	2	3	2
Slope (%)	12.7	2	0.7	2.3	2
Mean channel width (m)	2.5	4.6	2.5	4.5	2.6
Mean depth (m)	0.12	0.38	0.19	0.19	0.20
MAT (°C) [*]	0.9	1.8	10.2	15.4	25.4
M Jan T (°C) [*]	-9.7	-10.0	3.6	8.3	24.9
M July T (°C) [*]	12.3	15.2	17.4	23.9	25.2
MAP (mm) [*]	352	643	814	766	2519
Mean stream temperature during experiment (°C)	6.4	4.6	7.2	5.7	24.4
Temperature range during experiment (°C)	0.6-14.2	0.1-9.9	5.5-9.3	1.9-11.3	23.5-25.4
Conductivity (µS/cm)	122	19	233	190	19
pH	7.3	7.1	6.8	7.9	7.7
Dissolved oxygen (mg/L)	9.3	11.0	11.3	11.4	n.d.
Alkalinity (mmol/L)	0.78	0.55	0.33	1.33	0.23
N-NO ₃ ⁻ (mg/L)	0.02	0.27	7.45	0.18	<0.02
N-NH ₄ ⁺ (µg/L)	n.d.	<20	154	20	<5
SRP [†] (µg/L)	<2	18	16	18	<1
Terrestrial Locations	Subarctic	Boreal	Temperate	Mediterranean	Tropical
Site	Abisko, Sweden	Krycklan, Sweden	Leuvenemse Bos, The Netherlands	Barroubio, France	Paracou, French Guiana
Coordinates	68°21' N, 18°49' E	64°14' N, 19°50' E	52°18' N, 5°41' E	43°23' N, 2°51' E	5°18' N, 52°55' W
Dominant tree cover	<i>Betula pubescens</i>	<i>Betula pubescens</i> , <i>Picea abies</i>	<i>Fagus sylvatica</i>	<i>Quercus ilex</i> , <i>Fraxinus</i> <i>angustifolia</i>	~140 different canopy species
Leaf area index (LAI)	1.71 ± 0.15 (tree) 3.28 ± 0.29 (shrub)	6.5 [‡]	3.85 ± 0.12	3.66 ± 0.02	5.72 ± 0.25
MAT (°C) [*]	0.9	1.8	10.7	14.6	25.4
M Jan T (°C) [*]	-9.7	-10.0	4.3	7.4	24.9
M July T (°C) [*]	12.3	15.2	18.0	23.3	25.2
MAP (mm) [*]	352	643	911	670	2519
Soil pH	4.63 ± 0.27	4.70 ± 0.12	3.29 ± 0.06	6.56 ± 0.23	4.80 ± 0.04
Soil C _{mic} (µg g ⁻¹) [§]	4602 ± 2870	1266 ± 589	2937 ± 820	439 ± 87	648 ± 37
Soil C (%) [§]	27.36 ± 0.60	3.63 ± 0.06	25.21 ± 1.03	4.57 ± 0.72	2.84 ± 0.30
Soil N (%) [§]	0.95 ± 0.01	0.19 ± 0.00	1.14 ± 0.03	0.39 ± 0.04	0.19 ± 0.01

* Means were calculated based on 10-year records between 1998 and 2008 from the closest possible meteorological station.

[†] Soluble reactive phosphorus ≈ ortho-phosphate.

[‡] Data courtesy of K. Bishop and P.-E. Mellander.

[§] Soil microbial biomass (C_{mic}), soil C and soil N are expressed on a dry mass basis.

Extended Data Table 2 | Plant functional types, species identity and litter quality traits

Location	Plant Functional Type			
	N-fixing	Deciduous, Rapidly decomposing	Deciduous, Slow-decomposing	Evergreen
Subarctic	<i>Alnus incana</i> (L.) Moench	<i>Sorbus aucuparia</i> L.	<i>Populus tremula</i> L.	<i>Vaccinium vitis-idaea</i> L.
Boreal	<i>Alnus incana</i> (L.) Moench	<i>Prunus padus</i> L.	<i>Betula pubescens</i> Ehrh.	<i>Rhododendron tomentosum</i> Harmaja
Temperate	<i>Alnus glutinosa</i> (L.) Gaertn.	<i>Salix cinerea</i> L.	<i>Fagus sylvatica</i> L.	<i>Ilex aquifolium</i> L.
Mediterranean	<i>Alnus glutinosa</i> (L.) Gaertn.	<i>Fraxinus angustifolia</i> Vahl	<i>Pistacia terebinthus</i> L.	<i>Quercus ilex</i> L.
Tropical	<i>Diplotropis purpurea</i> (Rich.) Amshoff	<i>Qualea rosea</i> Aubl.	<i>Vochysia densiflora</i> Spruce ex Warm	<i>Eperua falcata</i> Aubl.
Quality trait ^a				
N	2.15 ± 0.28	0.97 ± 0.11	0.73 ± 0.10	1.05 ± 0.18
Lignin	17.2 ± 3.6	12.0 ± 2.9	16.5 ± 2.5	22.4 ± 4.1
Water saturation	238 ± 37	332 ± 182	222 ± 114	123 ± 28
Condensed tannins	1.48 ± 0.84	1.76 ± 0.54	2.68 ± 0.44	1.58 ± 0.35

Leaf litter was sampled from location-specific native tree species corresponding to four functional types (top) varying in quality traits associated with decomposition (bottom).

^aAll data are shown as percentage dry mass (mean ± s.e.m., $n = 5$). Methods are described in detail in ref.31.

Extended Data Table 3 | Results of analyses of variance testing for the net diversity effect (NE), complementarity effect (CE) and selection effect (SE) on C loss (top) and N loss (bottom) from decomposing leaf litter*

Source of variation	Net diversity			Complementarity			Selection		
	DF _{NE}	F _{NE}	P _{NE}	DF _{CE}	F _{CE}	P _{CE}	DF _{SE}	F _{SE}	P _{SE}
Effect on C loss									
N-fixers (NF)	1	3.53	0.06	1	0.64	0.42	1	0.06	0.80
Rapid decomposer (RD)	1	0.41	0.52	1	0.51	0.48	1	0.05	0.82
Slow decomposer (SD)	1	0.28	0.59	1	0.17	0.68	1	0.11	0.75
Evergreen (EG)	1	0.02	0.90	1	0.61	0.43	1	5.60	<0.05
NF × RD	1	5.25	<0.05	1	2.35	0.13	1	7.95	<0.01
RD × EG	1	0.63	0.43	1	0.00	0.99	1	7.64	<0.01
Decomposer comm. (DC)	2	0.40	0.67	2	1.11	0.33	2	3.29	<0.05
NF × DC	2	4.40	<0.05	2	1.99	0.14	2	2.60	0.07
RD × DC	2	5.09	<0.01	2	3.09	<0.05	2	3.59	<0.05
SD × DC	2	5.57	<0.01	2	4.43	<0.05	2	0.69	0.50
NF × SD × DC	3	0.41	0.74	3	0.28	0.83	3	5.08	<0.01
Location [†]	4	11.68	<0.001	4	11.08	<0.001	4	52.60	<0.001
NF × Location	4	1.98	0.10	4	4.12	<0.01	4	5.41	<0.001
RD × Location	4	3.58	<0.01	4	2.55	<0.05	4	4.11	<0.01
SD × Location	4	0.52	0.72	4	2.25	0.06	4	11.46	<0.001
EG × Location	4	1.51	0.20	4	2.17	0.07	4	13.63	<0.001
NF × RD × Location	4	1.11	0.35	4	2.15	0.07	4	3.02	<0.05
RD × EG × Location	4	1.73	0.14	4	1.67	0.15	4	3.08	<0.05
DC × Location	8	4.93	<0.001	8	4.36	<0.001	8	7.91	<0.001
NF × DC × Location	8	4.16	<0.001	8	2.28	<0.05	8	2.36	<0.05
RD × DC × Location	8	2.26	<0.05	8	1.54	0.14	8	1.57	0.13
SD × DC × Location	8	0.74	0.65	8	1.71	0.09	8	4.43	<0.001
NF × SD × DC × Location	12	0.91	0.53	12	1.12	0.34	12	3.26	<0.001
RD × SD × DC × Location	15	0.79	0.69	15	0.83	0.65	15	1.97	<0.05
Ecosystem type (ET) [†]	1	10.9	<0.001	1	15.8	<0.001	1	5.23	<0.05
RD × ET	1	0.45	0.50	1	0.57	0.45	1	22.17	<0.001
SD × ET	1	5.71	<0.05	1	1.99	0.16	1	5.42	<0.05
EG × ET	1	1.70	0.19	1	5.17	<0.05	1	9.74	<0.01
RD × SD × ET	1	0.89	0.35	1	1.52	0.22	1	5.42	<0.05
ET × Location [†]	4	11.12	<0.001	4	15.29	<0.001	4	8.58	<0.001
DC × ET	2	1.18	0.31	2	1.92	0.15	2	4.59	<0.05
DC × Location × ET	8	4.02	<0.001	8	4.30	<0.001	8	5.81	<0.001
NF × DC × Location × ET	15	2.09	<0.01	15	1.53	0.09	15	0.98	0.46
RD × DC × Location × ET	14	1.36	0.16	14	1.30	0.20	14	1.90	<0.05
Block	40	5.43	<0.001	40	5.38	<0.001	40	2.10	<0.001
Residuals	1437			1437			1437		
Effect on N loss									
N-fixers (NF)	1	0.63	0.43	1	0.02	0.88	1	0.22	0.63
Rapid decomposer (RD)	1	2.97	0.09	1	0.59	0.44	1	3.52	0.06
Slow decomposer (SD)	1	12.62	<0.001	1	7.93	<0.01	1	16.20	<0.001
Evergreen (EG)	1	0.04	0.83	1	1.44	0.23	1	1.27	0.26
SD × EG	1	0.20	0.65	1	8.67	<0.01	1	3.78	<0.05
Decomposer comm. (DC)	2	1.22	0.30	2	5.75	<0.01	2	3.39	<0.05
NF × DC	2	1.30	0.27	2	4.27	<0.05	2	0.76	0.47
RD × DC	2	4.23	<0.05	2	9.06	<0.001	2	5.90	<0.01
SD × DC	2	25.25	<0.001	2	11.72	<0.001	2	23.30	<0.001
Location [†]	4	6.28	<0.001	4	5.32	<0.001	4	7.01	<0.001
NF × Location	4	5.33	<0.001	4	0.68	0.61	4	1.93	0.10
SD × Location	4	5.87	<0.001	4	2.56	<0.01	4	5.36	<0.001
EG × Location	4	3.80	<0.01	4	1.84	0.12	4	3.22	<0.05
DC × Location	8	11.30	<0.001	8	14.03	<0.001	8	15.08	<0.001
NF × DC × Location	8	3.18	<0.01	8	3.18	<0.01	8	2.44	<0.05
RD × DC × Location	12	3.48	<0.001	12	3.68	<0.001	12	3.92	<0.001
SD × DC × Location	8	2.56	<0.01	8	6.93	<0.001	8	6.68	<0.001
Ecosystem type (ET) [†]	1	61.29	<0.001	1	49.46	<0.001	1	82.66	<0.001
ET × Location [†]	4	16.42	<0.001	4	19.68	<0.001	4	20.08	<0.001
NF × ET	1	10.46	<0.01	1	1.31	0.25	1	9.47	<0.01
NF × Location × ET	4	8.33	<0.001	4	4.12	<0.01	4	6.93	<0.001
NF × RD × ET	2	6.37	<0.01	2	2.10	0.12	2	3.89	<0.05
SD × ET	1	4.54	<0.05	1	30.09	<0.001	1	15.92	<0.001
SD × Location × ET	4	10.69	<0.001	4	17.78	<0.001	4	17.21	<0.001
NF × SD × Location × ET	10	2.19	<0.05	10	5.03	<0.001	10	5.21	<0.001
EG × Location × ET	5	2.75	<0.05	5	3.42	<0.01	5	2.40	<0.05
DC × ET	2	15.11	<0.001	2	21.04	<0.001	2	19.83	<0.001
RD × DC × ET	2	0.17	0.85	2	7.70	<0.001	2	3.63	<0.05
DC × Location × ET	8	5.96	<0.001	8	7.90	<0.001	8	6.70	<0.001
NF × DC × Location × ET	10	2.81	<0.01	10	1.59	0.11	10	1.01	0.43
SD × DC × ET	2	5.26	<0.01	2	8.19	<0.001	2	9.11	<0.001
SD × DC × Location × ET	8	4.17	<0.001	8	4.52	<0.001	8	5.78	<0.001
Block	40	2.55	<0.001	40	2.44	<0.001	40	1.80	<0.01
Residuals	1460			1460			1460		

* Interaction terms omitted from the final model are not significant for any of the three response variables.

† Location, ecosystem type and their interaction were tested against the block rather than against the residual.

Extended Data Table 4 | Characteristics of stream macroinvertebrate communities at the five tested locations*

Location	Detritivores		Detritivore community composition [†]				Predators		
	Density (# m ⁻²)	Taxon richness	Plecoptera (%)	Trichoptera (%)	Gammarus (%)	Dominant taxa	Density (# m ⁻²)	Taxon richness	Dominant taxa
Subarctic	46.2 ±12.8	5	95.7 ±4.8	4.3 ±4.8	-	<i>Leuctra digitata</i> , <i>Capnopsis schilleri</i>	40.0 ±20.4	7	<i>Dicranota</i> , <i>Rhyacophila</i> , <i>Diura nanseni</i>
Boreal	267.3 ±258.5	8	96.8 ±1.3	3.2 ±1.2	-	<i>Protonemura meyeri</i> , <i>Nemoura flexuosa</i> , <i>Leuctra digitata</i>	46.6 ±30.7	8	<i>Isoperla</i> , <i>Rhyacophila nubila</i>
Temperate	332.3 ±62.6	13	4.3 ±5.8	10.2 ±16.1	82.3 ±13.7	<i>Gammarus pulex</i> , <i>Chaetopteryx villosa</i> , <i>Sericostoma personatum</i>	130.7 ±47.4	13	<i>Dugesia gonocephala</i> , <i>Plectrocnemia conspersa</i>
Mediterranean [‡]	432.1 ±330.6	10	62.9 ±26.2	36.1 ±26.6	1.0 ±2.2	Leuctridae, Taeniopterygidae, Nemouridae, Limnephilidae	418.3 ±239.1	11	Perlodidae, Athericidae, Limoniidae, Rhyacophilidae
Tropical	8.6 ±14.7	3	-	21.4 ±39.3	-	<i>Phylloicus</i> , Pyralidae, Tipulidae	12.3 ±22.8	4	Perlidae, Calopterygidae

*The mean density and total taxon richness of detritivores, their main invertebrate predators, and the mean proportion of Plecoptera, Trichoptera and Gammarus as a percentage of total detritivore abundance (mean ± s.d.) are shown. All samples were collected using a 500-µm mesh net at the same time of year as the main experiment (although in different years in some cases). Specific sampling protocols differed between locations, with the density standardized to the number of individuals per metre squared. For the subarctic site, six replicate kick samples were taken, each from an area of 1 m × 0.35 m for 1 min, during September 2006. Identification was mostly to the species level³⁴. For the boreal site, four replicate Surber samples per year were taken for three years, during October 2010–2012, with a quadrat size of 0.25 m × 0.5 m. Identification was mostly to the species level (B.G.M. and P.-O. Hoffsten, unpublished observations). For the temperate site, five replicate sweep net samples were taken, each from an area of 0.3 m × 5 m, in October 1992. Identification was mostly to the species level³⁵. For the Mediterranean site, five replicate Surber samples were taken, with a quadrat size of 0.33 m × 0.31 m, in January 2014. Identification was mostly to the family level (E.C. and S. Lamothe, unpublished observations). For the tropical site, ten replicate natural leaf packs (first-sized handfuls of leaves picked from the stream bed) were taken from each of seven streams in May 2007. Abundances per leaf pack were converted to densities based on standardized visual estimates of stream-bed litter cover. Identification was mostly to the family level (A.B., M. Schindler, M. S. Moretti and M.O.G., unpublished observations).

[†]Detritivore community composition data do not sum to 100% at all locations, owing to the presence of other dipteran (Tipulidae), lepidopteran (Pyralidae) and crustacean (Asellidae) shredders at the temperate site, and tipulid and pyralid shredders at the tropical site.

[‡]The caddisfly *Micrasema* (Brachycentridae) was common at the Mediterranean site but was small and was not regarded as a shredder.

Extended Data Table 5 | Full model output of the relative contributions of variance associated with diversity and sites to explain C and N loss

Source of variation	DF	C loss			N loss		
		%SS	F	P	%SS	F	P
Variation associated with diversity							
Litter community							
Litter richness [†]	1	0	0.01	0.93	1.0	63.7	***
Litter diversity [†]	2	0.1	2.61	0.07	0.3	8.29	***
Variation associated with functional type (FT) composition [‡]	11	10.0	77.5	***	8.6	50.9	***
Decomposer community (DC)	2	5.8	247	***	4.4	142	***
Litter community x DC							
Richness x DC	2	0	0.37	0.69	0.1	3.45	*
Diversity x DC	4	0	0.70	0.59	0.0	0.07	0.99
FT composition x DC	22	0.8	2.96	***	1.2	3.61	***
Variation associated with sites							
Location	4	12.1	52.8	***	16.9	66.5	***
Richness x Location	4	0.1	2.07	0.08	0.3	5.06	***
Diversity x Location	8	0.2	2.35	*	0.3	2.37	*
FT compositions x Location	44	15.8	30.6	***	10.4	15.4	***
DC x Location	8	8.6	92.1	***	7.5	61.3	***
Richness x DC x Location	8	0	0.51	0.85	0.1	0.53	0.83
Diversity x DC x Location	16	0.3	1.36	0.15	0.2	0.84	0.64
FT composition x DC x Location	88	1.6	1.56	***	1.7	1.28	*
Ecosystem	1	6.4	113	***	1.4	22.3	***
Location x Ecosystem	4	6.3	27.7	***	8.0	31.6	***
Richness x Ecosystem	1	0.1	5.79	*	0	1.02	0.31
Richness x Location x Ecosystem	4	0.1	1.25	0.29	0.2	3.21	*
Diversity x Ecosystem	2	0	0.62	0.54	0	1.27	0.28
Diversity x Location x Ecosystem	8	0.2	2.37	*	0.1	1.11	0.35
FT Composition x Ecosystem	11	1.0	7.94	***	1.4	8.30	***
FT Composition x Location x Ecosystem	44	1.7	3.30	***	1.9	2.81	***
DC x Ecosystem	2	1.9	79.7	***	0.9	27.7	***
DC x Location x Ecosystem	8	2.4	25.7	***	1.9	15.9	***
Richness x DC x Ecosystem	2	0	0.36	0.70	0	0.60	0.55
Richness x DC x Location x Ecosystem	8	0.1	0.70	0.69	0	0.33	0.96
Diversity x DC x Ecosystem	4	0	0.35	0.84	0.1	1.40	0.23
Diversity x DC x Location x Ecosystem	16	0.1	0.38	0.99	0.2	0.79	0.70
FT Composition x DC x Ecosystem	22	0.3	1.14	0.30	0.4	1.08	0.36
FT Composition x DC x Location x Ecosystem	88	1.2	1.17	0.13	1.2	0.92	0.69
Block (within location x ecosystem)	40	2.3	4.87	***	2.5	4.13	***
Total variance explained by the model	489	79.6	13.9	***	73.3	9.76	***
Residuals	1739	20.4			26.7		
Total	2228	100.0			100.0		

Variance associated with diversity and sites is expressed in percentage sums of squares (% SS); levels of significance are * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

[†] Plant species were selected to represent the same four functional types (FTs) at each location (N-fixing, evergreen, rapidly decomposing deciduous trees or shrubs, or slowly decomposing deciduous trees or shrubs). Litter FT richness (linear contrast) and litter diversity (factorial contrast) were fitted before litter FT compositions.

[‡] An alternative model omitting richness and testing in detail the litter FT compositions in a full factorial analysis with contrasts for FT presence/absence and interactions is presented in Extended Data Table 7. That model highlights the importance of the interaction between the litter of the N-fixing FT and the rapidly decomposing FT, hinting at a N-transfer mechanism.

Extended Data Table 6 | Characteristics of soil fauna communities at the five tested locations*

Location	Mesofauna [†]						
	Microbi-detritivores		Dominant taxa [‡]		Predators		Dominant taxa [‡]
	Density (# m ⁻²)	Taxon richness [†]	Collembola (%)	Acari (%)	Density (# m ⁻²)	Taxon richness [†]	
Subarctic	4068 ±2254	8	46.9 ±13.0	53.1 ±13.0	506 ±209	1	Gamasida
Boreal	9824 ±3750	10	61.9 ±14.9	38.1 ±14.9	690 ±356	1	Gamasida
Temperate	17020 ±6846	11	75.0 ±8.3	25.0 ±8.3	711 ±347	2	Gamasida (73%) Diplura (27%)
Mediterranean	9222 ±6514	10	21.9 ±6.9	78.1 ±6.9	1386 ±598	2	Gamasida (64%) Diplura (36%)
Tropical	641 ±134	6	3.9 ±3.7	96.0 ±3.7	276 ±106	3	Gamasida (47%) Diplura (53%)
Location	Macrofauna [†]						
	Detritivores		Dominant taxa [‡]		Predators		Dominant taxa [‡]
	Density (# m ⁻²)	Taxon richness [†]			Density (# m ⁻²)	Taxon richness [†]	
Subarctic	595 ±249	10	Diptera (68%) Gastropoda (25%)		588 ±197	9	Araneidae (41%) Diptera (35%)
Boreal	333 ±185	9	Diptera (82%) Annelida (13%)		1824 ±1674	7	Diptera (73%) Coleoptera (20%)
Temperate	81 ±59	6	Diplopoda (60%) Gastropoda (23%)		3640 ±1534	8	Chilopoda (88%)
Mediterranean	501 ±219	10	Isopoda (43%) Diptera (38%)		2244 ±2021	16	Chilopoda (51%) Hymenoptera (30%)
Tropical	1750 ± 3115	10	Isoptera (34%) Diptera (26%)		1813 ±1515	13	Hymenoptera (59%) Pseudoscorpionida (32%)

* The mean density, the total taxon richness and the proportion of dominant taxa as a percentage of total community abundance are shown (mean ± s.d.). Communities are divided into mesofauna and macrofauna, reflecting an increase in body size that relates to mesh size differences in the field microcosms.

† Taxon richness is based on the number of observed families.

‡ Dominant taxa data are based on a lower taxonomic resolution than taxon richness, mainly order or class level. The community composition data do not always sum to 100% at all locations, owing to the presence of other taxa.

§ All samples were collected at the end of the growing season in 2008 (subarctic and boreal, late September; temperate and Mediterranean, October; and tropical, early December). Eight Kempson cores (21-cm diameter) and eight MacFayden cores (5-cm diameter) were taken at each field site. The reported data are based on extraction of the whole soil core (9-cm height), including the litter layer. Soil arthropods were extracted, counted and identified to the highest possible taxonomic level (families) (O.B. and S.S., unpublished observations).

Extended Data Table 7 | Analysis of variance testing for effects on total litter C loss (top) and N loss (bottom)

Source of variation	DF	SS	F	% Sum of Squares
% C Loss				
Rapid decomposer (RD)	1	16580	290.7	3.9
Slow decomposer (SD)	1	12068	211.6	2.9
N-fixer (NF)	1	3146	55.2	0.7
Evergreen (EG)	1	3576	62.7	0.9
RD × NF	1	3532	61.9	0.8
RD × EG	1	747	13.1	0.2
SD × EG	1	1242	21.8	0.3
RD × SD	1	223	3.9	0.1
Decomposer community [*]	1	23404	410.3	5.6
Decomposer community × NF	1	907	15.9	0.2
Decomposer community × SD	1	526	9.2	0.1
Location [†]	4	50698	51.5	12.1
Location × EG	4	26247	115.0	6.2
Location × RD	4	13398	58.7	3.2
Location × SD	4	7784	34.1	1.9
Location × NF	4	8818	38.7	2.1
Location × RD × EG	4	1272	5.6	0.3
Location × RD × SD	4	1686	7.4	0.4
Location × SD × EG	4	976	4.3	0.2
Location × Decomposer community	4	31306	137.2	7.5
Location × Decomposer community × SD	4	1144	5.0	0.3
Ecosystem type (ET) [†]	1	27126	110.3	6.5
ET × Location [†]	4	26703	27.1	6.4
ET × Location × RD	5	1330	4.7	0.3
ET × Location × SD	5	1860	6.5	0.4
ET × NF	1	446	7.8	0.1
ET × Location × NF	4	554	2.4	0.1
ET × EG	1	3450	60.5	0.8
ET × Location × EG	4	2318	10.2	0.6
ET × RD × SD	1	262	4.6	0.1
ET × Decomposer community	1	7761	136.1	1.8
ET × Location × Decomposer community	4	8908	39.0	2.1
Block	40	9836	4.3	2.3
Residuals	2106	120133		28.6
% N Loss				
Rapid decomposer (RD)	1	16570	87.7	1.4
Slow decomposer (SD)	1	31927	169.0	2.7
N-fixer (NF)	1	40390	213.7	3.5
Evergreen (EG)	1	3021	16.0	0.3
RD × NF	1	13245	70.1	1.1
RD × EG	1	2697	14.3	0.2
SD × NF	1	971	5.1	0.1
NF × EG	1	2438	12.9	0.2
RD × NF × EG	1	2947	15.6	0.3
Decomposer community [*]	1	48999	259.3	4.2
Decomposer community × NF	1	5281	28.0	0.5
Decomposer community × SD	1	1080	5.7	0.1
Decomposer community × EG	1	771	4.1	0.1
Decomposer community × NF × SD	1	1071	5.7	0.1
Decomposer community × NF × EG	1	2051	10.9	0.2
Location [†]	4	195881	66.7	16.8
Location × Decomposer community	4	75781	100.3	6.5
Location × NF	4	46245	61.2	4.0
Location × RD	4	30727	40.7	2.6
Location × EG	4	20789	27.5	1.8
Location × SD	4	4815	6.4	0.4
Location × NF × EG	4	10603	14.0	0.9
Location × RD × EG	4	3757	5.0	0.3
Location × NF × RD	4	3346	4.4	0.3
Location × Decomposer community × EG	4	2632	3.5	0.2
Ecosystem type (ET) [†]	1	16565	22.6	1.4
ET × Location [†]	4	92914	31.7	8.0
ET × RD	1	5845	30.9	0.5
ET × NF	1	1207	6.4	0.1
ET × Location × NF	4	9700	12.8	0.8
ET × EG	1	4173	22.1	0.4
ET × Location × EG	4	3870	5.1	0.3
ET × NF × RD	1	2554	13.5	0.2
ET × Location × RD × NF	4	1908	2.5	0.2
ET × NF × EG	1	1419	7.5	0.1
ET × Decomposer community	1	9015	47.7	0.8
ET × Location × Decomposer community	4	17968	23.8	1.5
Block	40	29349	3.9	2.5
Residuals	2106	397977		34.2

All terms included in the final model shown are significant at $P < 0.05$.

*The decomposer community was fitted as a log-linear contrast and not a factorial contrast (a factorial contrast is shown in Table 1).

†The location and ecosystem type and their interaction were tested against the block rather than against the residual.

Extended Data Table 8 | Analysis of variance testing for the proportional change in total litter N content

Source of variation	DF	SS	F	P
Mixture (decomposing alone or in presence of the other functional type)	1	0.002	0.057	0.81
Functional type (N-fixing or rapidly decomposing)	1	0.650	15.8	<0.001
Mixture × Functional type [*]	1	0.290	7.05	<0.01
Location [†]	4	2.053	11.9	<0.05
Ecosystem type (ET) [†]	1	0.818	18.9	<0.05
ET × Location [†]	4	2.277	13.1	<0.05
Block	4	0.173	1.05	0.38
Residuals	574	23.61		

The test compares specific two-species combinations, including the particular functional types of N-fixing and rapidly decomposing plants, to their respective single-species treatment across location and ecosystem type.

* A significant difference in the mixture × functional type interaction is taken as an indication of N transfer between litter species.

† Location and ecosystem type and their interaction were also included as factors in this analysis and were tested against the block rather than the residuals.

Extended Data Table 9 | Experimental duration and richness of naturally occurring local litter species in terrestrial and aquatic ecosystems at each of five widely dispersed locations

Ecosystem	Location	Incubation start	Incubation end	Duration (d) [*]	Natural litter species richness [†]
Terrestrial	Subarctic	22/ 09/ 06	5/ 09/ 08	714	6
	Boreal	20/ 10/ 06	12/ 05/ 08	570	4
	Temperate	17/ 01/ 07	5/ 03/ 08	414	1
	Mediterranean	15/ 11/ 07	19/ 10/ 08	338	4
	Tropical	20/ 11/ 06	25/ 08/ 07	278	7
Aquatic	Subarctic	27/ 07/ 07	11/ 10/ 07	76	6
	Boreal	11/ 09/ 07	13/ 11/ 07	63	4
	Temperate	18/ 10/ 07	13/ 12/ 07	56	4
	Mediterranean	16/ 11/ 07	3/ 01/ 08	48	11
	Tropical	9/ 05/ 07	14/ 08/ 07	97	10

^{*} Incubation dates differed across ecosystem types and locations to ensure that, at the time of sampling, 40–50% of the mass of the most rapidly decomposing litter remained, thus allowing comparisons at similar decomposition stages.

[†] The mean species richness counts of naturally occurring litter in five randomly sampled plots that were the size of microcosms (15-cm diameter) in each of the five experimental blocks.

Niche filling slows the diversification of Himalayan songbirds

Trevor D. Price¹, Daniel M. Hooper¹, Caitlyn D. Buchanan¹, Ulf S. Johansson^{1,2}, D. Thomas Tietze^{1,3}, Per Alström^{4,5}, Urban Olsson⁶, Mousumi Ghosh-Harihar⁷, Farah Ishtiaq⁷, Sandeep K. Gupta⁷, Jochen Martens⁸, Bettina Harr⁹, Pratap Singh⁷ & Dhananjai Mohan⁷

Speciation generally involves a three-step process—range expansion, range fragmentation and the development of reproductive isolation between spatially separated populations^{1,2}. Speciation relies on cycling through these three steps and each may limit the rate at which new species form^{1,3}. We estimate phylogenetic relationships among all Himalayan songbirds to ask whether the development of reproductive isolation and ecological competition, both factors that limit range expansions⁴, set an ultimate limit on speciation. Based on a phylogeny for all 358 species distributed along the eastern elevational gradient, here we show that body size and shape differences evolved early in the radiation, with the elevational band occupied by a species evolving later. These results are consistent with competition for niche space limiting species accumulation⁵. Even the elevation dimension seems to be approaching ecological saturation, because the closest relatives both inside the assemblage and elsewhere in the Himalayas are on average separated by more than five million years, which is longer than it generally takes for reproductive isolation to be completed^{2,3,6}; also, elevational distributions are well explained by resource availability, notably the abundance of arthropods, and not by differences in diversification rates in different elevational zones. Our results imply that speciation rate is ultimately set by niche filling (that is, ecological competition for resources), rather than by the rate of acquisition of reproductive isolation.

Range expansions are a critical step in the speciation cycle: without them, allopatric and parapatric forms would have ranges of ever decreasing size, unlikely to be further fragmented by barriers⁷. The expansion of geographical range by one taxon inevitably brings it into sympatry with related taxa, which requires reproductive isolation between the forms^{1,8} and is generally thought to require ecological differences as well^{1,9}. In young adaptive radiations reproductive isolation and ecological divergence may be coupled^{10,11}, with consequent rapid cycling through the speciation cycle² (Extended Data Fig. 1). For example, young species of Darwin's ground finches (*Geospiza*) differ in beak and body size and coexist in sympatry by exploiting different-sized seeds. These beak and body size differences contribute to reproductive isolation, because they are used as cues in conspecific mate choice (premating isolation) and because intermediate-sized hybrids are at a disadvantage in some environments (postmating isolation)¹¹. As adaptive radiations proceed and environments fill with species, rates of ecological divergence slow^{10,12} and concomitantly the rate of evolution of reproductive isolation declines. Evidence from bird hybrid zones on continents implies that species may continue to interbreed even when separated by more than two million years^{2,3}, preventing mutual range expansions⁴ and delaying the speciation cycle³.

Here, we introduce a method of studying the causes of slowing speciation rates as species accumulate by considering all species in a sympatric continental community. We ask whether limits to range expansions into the community are attributable to an absence of reproductive isolation

or to ecological competition, by considering the age and pattern of diversification along different ecological dimensions, and by directly evaluating species distributions through field measurements of resources. To do this, we built a molecular-based phylogeny for all 461 Himalayan songbirds (Fig. 1). The songbirds (or oscines) are one of three suborders in the order Passeriformes, which is one of the 39 bird orders, but this one suborder contains more than 5,000 species, or over 45% of all the world's birds. Worldwide, they reach their maximum diversity in the eastern Himalayas, where about 358 species breed within our 10,000-km² study area (Fig. 1). We consider these species to be sympatric because they are within “cruising range” of one another⁸. They occupy a diversity of climates, from near-tropical to near-boreal, their body mass varies >200-fold (4.5 g to 950 g), their shape differences are extreme (for example, among species weighing about 30 g, beak length varies fivefold) and they include specialist nectarivores, frugivores, granivores, insectivores and aerial foragers, that is, they bear the hallmarks of an adaptive radiation. We assume that ecological differences are essential to coexistence. Previous work on one subgroup of similar species, the Old World leaf warblers (Phylloscopidae, 19 Himalayan species¹³), demonstrated the importance of ecological differences, with sympatric species differing along at least one dimension of feeding method, body size and elevation^{14,15}.

Biogeographical reconstructions relate many of the speciation events that led to the 358 species in our study area to barrier formation outside the Himalayas, including between the Indonesian islands¹⁶ and within China¹³. Thus, the east Himalayan community has been assembled largely by immigration from outside, with ecological differences either generated *in situ*, or before invasion. We emphasize relationships within a group of sympatric species, rather than within the entire songbird clade, for two reasons. First, many of the other species in the songbird clade (for example, Darwin's finches) have originated in response to ecological conditions elsewhere, so including them would require a complex integration of the timing of establishment and patterns of ecological diversification across the world. Second, many closely related allopatric species apparently occupy the same ecological niche in different places. For example, Päckert *et al.*¹⁷ studied 26 sister species pairs which contained at least one species in our study area; they found that in most cases (63%) the other sister was a similar, allopatric, replacement (see also Extended Data Fig. 2). A collection of closely allopatric replacements that occupy the same niche in different places form a superspecies¹. Phylogenetic relationships among sympatric species thus capture diversification patterns among superspecies and hence lineage splits that can potentially be limited by niche filling (Extended Data Figs 2 and 3).

Previous phylogenetic analyses of two subgroups of Himalayan songbirds, the Old World leaf warblers¹⁴ and separately, the Corvoidea¹⁸ (57 species) found that differences in elevation generally evolved more recently than differences in feeding method and body size. In Fig. 2 we show this to be the case for the entire 358 species assemblage. Consistent with niche filling along the body size and feeding habit dimensions,

¹Department of Ecology and Evolution, University of Chicago, Chicago, Illinois 60637, USA. ²Department of Zoology, Swedish Museum of Natural History, 10405 Stockholm, Sweden. ³Institute of Pharmacy and Molecular Biotechnology, University of Heidelberg, Im Neuenheimer Feld 364, 69120 Heidelberg, Germany. ⁴Key Laboratory of Zoological Systematics and Evolution, Institute of Zoology, Chinese Academy of Sciences, 1 Beichen West Road, Chaoyang District, Beijing 100101, China. ⁵Swedish Species Information Centre, Swedish University of Agricultural Sciences, Box 7007, 75007 Uppsala, Sweden. ⁶Systematics and Biodiversity, Department of Biology and Environmental Sciences, University of Gothenburg, 40530 Gothenburg, Sweden. ⁷Wildlife Institute of India, PO Box 18, Chandrabani, Dehradun 248001, India. ⁸Institute of Zoology, Johannes Gutenberg University, Mainz 55099, Germany. ⁹Max Planck Institute for Evolutionary Biology, August Thienemannstrasse 2, 24306 Plön, Germany.

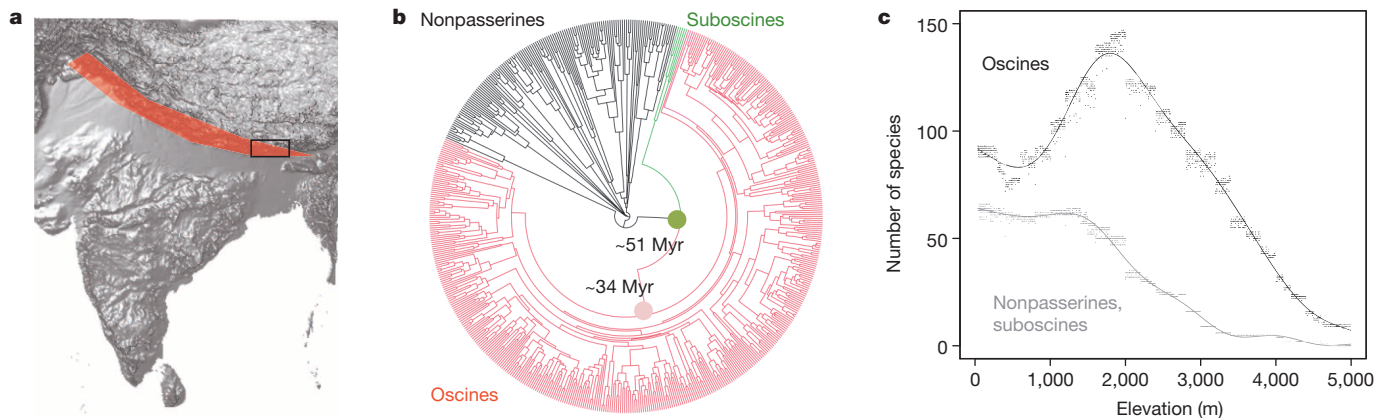


Figure 1 | Phylogeny and species distributions. **a**, Red shading indicates the area covered by the phylogeny and the rectangle denotes the study area. **b**, Estimate of phylogenetic relationships among all Himalayan songbirds (red), suboscines (a second suborder of the Passeriformes, green) and ten orders of nonpasserines (black), computed in BEAST²⁸ and dated using multiple fossils. 95% confidence limits on dates based on the Bayesian sample are

approximately ± 3 Myr. **c**, Local abundance of songbirds (oscines) and suboscines + nonpasserines in the study area as assessed by overlapping published range maps with elevational distributions. Lines are nonparametric curve fits. Each point represents the total number of species thought to be breeding in a $0.04^\circ \times 0.04^\circ$ square (about 2 km^2).

rates of diversification in size and shape have significantly slowed towards the present (Extended Data Figs 4 and 5) and clades dating back beyond ten million years (Myr) ago occupy restricted morphological space (for example, Fig. 2b). In contrast, differentiation along the elevation axis has occurred more recently, albeit still many millions of years ago (Fig. 2a, Extended Data Fig. 4). Along with these ecological patterns, lineage splitting in the 358 species phylogeny declines precipitously through time (Fig. 2, Extended Data Table 1). Strikingly, the average time since separation (\pm the standard error, throughout) of pairs of close relatives ('sister pairs' on the assemblage phylogeny) is estimated as $7.1 \text{ Myr} \pm 0.3$, $N = 116$ pairs (97.5% lower bound based on a sample of trees from the a posteriori distribution, $5.6 \text{ Myr} \pm 0.26$). These results are consistent with the idea that niche availability limits the establishment of species in the assemblage, with new ecological opportunity¹² along the elevation dimension appearing most recently, perhaps in association with documented climate change (and turnover in the mammalian fossil record) 6 to 10 Myr ago¹⁹.

Although patterns of evolution suggest a dynamic whereby niche-filling slows species accumulation, a failure of species to become established in the assemblage may instead reflect insufficient time to develop reproductive isolation between allopatric forms. However, a strong slowdown in lineage diversification is present even if the phylogeny is truncated at 3 Myr ago (Extended Data Table 1), which is a reasonable upper bound on the time it takes for reproductive isolation to be completed in nature²⁶; establishment of allopatric forms that are older than this should not be limited by reproductive isolation. Further, 85 Himalayan songbird species are found to the west of our study area. They are separated by an average of 7.5 ± 0.38 Myr from their closest relative in the study area, and 77 of these 85 species breed alongside this close relative somewhere in their western range (overlap of at least $1,500 \text{ km}^2$). Thus, the failure of these 77 species to expand their range into the study area cannot be attributed to incomplete reproductive isolation. We directly evaluated the niche-filling explanation by studying species richness patterns along the elevational gradient. This is the ecological axis associated with the

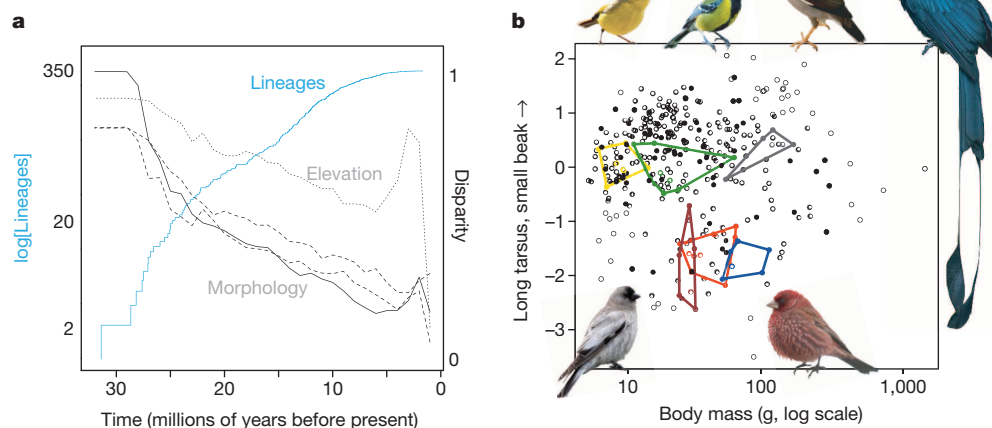


Figure 2 | Morphological evolution. **a**, Number of lineages at a given time slice through the phylogeny (blue solid line) and disparity measures for three morphological traits—mass (solid black line) and two shape measures (dashed black lines)—and elevation (taken as the mid-point of range, dotted black line) for the east Himalayan songbirds. Disparity is a measure of the fraction of the total variance that is within the clades subtended by any timeline²⁹. All three disparity measures deviate significantly from a Brownian motion model

(Extended Data Figs 4 and 5). **b**, First shape component versus mass. Polygons enclose morphologies for those six lineages at 11.5 Myr ago that subtend at least six species in the study area. One species from each clade is indicated (clockwise from top left *Phylloscopus xanthoschistos*, *Parus monticolus*, *Acridotheres fuscus*, *Dicrurus paradiseus*, *Carpodacus rubicilla* and *Leucosticte brandti*). The solid points are individual species with no shared relative dating to 11.5 Myr ago. Separate analyses for the major subclades are in Supplementary Table 1.

most recent divergences (Fig. 2) and hence a priori is the dimension most likely to contain additional opportunities for species to differentiate along. Songbirds peak in abundance at an altitude of about 1,800 m in the east Himalayas (Fig. 1). We ask why this should be.

The most popular model that incorporates the slow development of reproductive isolation as a potential explanation for patterns of species richness along elevational gradients considers that each climatic zone independently accumulates species, and different zones accumulate species at different rates²⁰. In this model, species and clades present in one zone remain confined to that zone because they are poorly adapted to alternative climates (climatic niche conservatism²⁰). In contrast to the predictions of this model, we found that the average age of separation of species in the assemblage declines monotonically with elevation, rather than being lowest in the most species-rich elevational band (Fig. 3a, Extended Data Fig. 6). Further, even at the high elevations, the plot of lineage diversity versus time shows an exceptionally strong slowdown, with close relatives separated by millions of years (Extended Data Fig. 6), suggesting that plenty of time has been available to accumulate additional species in this zone. Finally, climatic niche conservatism along the elevational gradient is weak, with close relatives often found in different zones (see the elevation disparity curve in Fig. 2 and ref. 21). Therefore, there is little support in these data for climatic niche conservatism: newly formed species should be able to transition rapidly between climatic zones in the absence of any other limits.

In models of niche filling, more species are found where resources are either more diverse or more abundant²². We tested the resource diversity hypothesis as an explanation for species distribution along the elevational gradient by using morphological diversity as a surrogate²³. We found that morphological diversity increases from the lowest to intermediate elevations, as may be expected given the increase in species numbers. However, morphological diversity increases still further to the highest elevations, despite the relatively few species at these elevations (Fig. 3b, Extended Data Fig. 7). Exploitation of a relatively high diversity of resources at high elevations has been attributed to a release from competition from the nonpasserines¹⁸, which predominate at lower elevations (Fig. 1c) and may also reflect the presence of additional open country habitats. Whatever the underlying causes, the monotonic increase in the apparent diversity of resources exploited does not match the mid-elevation peak in species richness.

Finally, in a hypothesis of resource abundance a “poor environment supplies too meager a resource base for its would-be rarest species, and they become extinct” (page 56 of ref. 24). To investigate this, over six years, we established eighteen 5-hectare grids in forested habitats at all altitudes across the study area (Extended Data Figs 8 and 9). We censused breeding birds, classified them according to feeding habit, measured

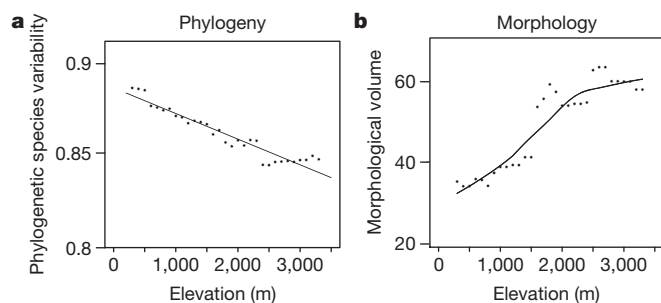


Figure 3 | Phylogenetic and morphological diversity along the elevational gradient. **a**, Estimate of phylogenetic relatedness³⁰ for all songbirds in a given elevational band; the line is the least-squares regression. On the scale, a value of 1 would imply a star phylogeny (that is, all species have independent history), whereas a value close to 0 implies that many species are closely related to each other. **b**, Volume of the convex hull for three morphological dimensions (see Extended Data Fig. 5a), as occupied by all songbirds in a given elevational band (the line is a nonparametric line fit; the outlier *Xiphirhynchus superciliosus* is omitted from this plot; its inclusion greatly steepens the slope).

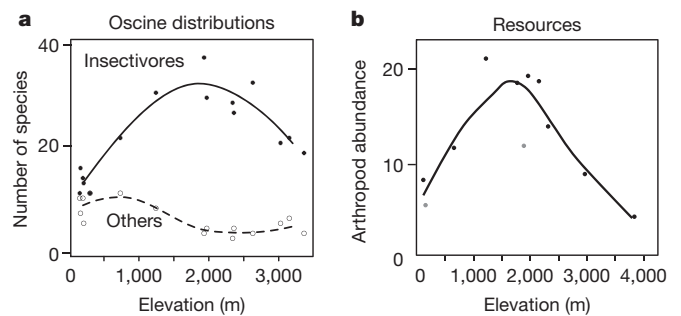


Figure 4 | Species richness and resource abundance. **a**, Number of breeding species censused on fifteen 5-hectare plots. **b**, Arthropod densities (from ref. 15, plus additional data (grey points)). Lines are nonparametric curve fits.

vegetation parameters and, for a subset of grids, we also measured arthropod densities. The mid-elevation peak in these 5-hectare plots (Fig. 4a) is very similar to that found at the larger scale (Fig. 1c). We found that this peak is entirely attributable to insectivorous species (Fig. 4a) and strikingly matches arthropod abundance (Fig. 4b). Arthropod abundance correlates with the total number of insectivorous individual birds on the 5-hectare grid (from 44 to 218; Pearson's correlation $r = 0.76$, the correlation with insectivorous bird biomass is $r = 0.75$) and the more insectivorous birds on the 5-hectare grid, the more insectivorous species were present ($r = 0.93$). A mid-elevation peak in arthropod abundance is surprising, given that both rainfall and temperature monotonically decline with elevation (Extended Data Fig. 8). However, mid-elevations are the lowest elevations that regularly freeze in winter and are therefore likely to experience a large leaf and arthropod flush in the spring. In addition, the lowest elevations contain many other insectivorous animals, potentially in competition with the songbirds. These include nonpasserine and suboscine birds (Fig. 1), an insectivorous ant (*Oecophylla smaragdina*, which is abundant at low elevations but absent at mid-elevations, personal observation of D.M.H. in 2011) and probably other poikilotherms.

In this continental environment, where species have had time to accumulate and dispersal barriers are relatively weak, both the appearance of new species and morphological diversification have substantially slowed and species distributions are well explained by abundance of resources. Rabosky and Matute²⁵ have demonstrated that, across clades, rates of evolution of reproductive isolation are generally decoupled from diversification rates. Our results imply that the ultimate limit on diversification is best explained, not by a slow rate of accumulation of reproductive isolation, but by the failure of species to expand ranges into new localities, which we attribute to competitive interactions. Because range expansions are essential to most ongoing speciation events³, local interactions ultimately determine regional speciation rates, rather than regional speciation rates setting local diversity, as in many non-equilibrium models²⁶. These results contrast with those from young environments, such as young archipelagos¹¹—which provide some of the best evidence for ongoing adaptive radiations—and with those from dispersal-limited, extinction-prone environments, notably islands—which provide some of the best evidence for bird communities below saturation²⁷.

METHODS SUMMARY

Tree construction. We obtained mitochondrial sequence data for all but one passerine species and nuclear data for 89%. We estimated a tree for the passerines and separately for 133 nonpasserines (all but 15 Himalayan species) from the following orders—the Piciformes, Bucerotiformes, Upupiformes, Trogoniformes, Coraciiformes, Cuculiformes, Psittaciformes, Apodiformes, Strigiformes and Columbiformes. For the primary analyses we constructed an ultrametric tree in BEAST²⁸, dated using multiple fossils and biogeographic dates, all set with uniform priors.

Morphology. We measured two or more specimens of each species in museums, drawing on males from within the study area whenever possible. We measured beak depth, beak length, beak width, tarsus length, and wing length. Mass and elevational distributions were based on the primary literature and our own fieldwork. In the main analyses we used $\ln(\text{mass})$ and the first two principal components extracted

from the correlation matrix of residuals of regressions of $\ln(\text{morphological traits})$ on $\ln(\text{mass})$, all standardized to have unit variance (alternative approaches give similar results; see Extended Data Fig. 5).

Fieldwork. T.D.P. and D.M., plus up to two other experienced workers, independently censused birds in eighteen 5-hectare grids spanning the elevational gradient. In each grid two mornings were spent territory-mapping males, largely following the British Trust for Ornithology's common bird census protocols (<http://www.bto.org>) and assuming that the number of birds was twice the number of censused males.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 December 2013; accepted 24 March 2014.

Published online 30 April 2014.

- Mayr, E. & Diamond, J. M. *The Birds of Northern Melanesia: Speciation, Ecology, and Biogeography* (Oxford Univ. Press, 2001).
- Price, T. *Speciation in Birds* (Roberts, 2008).
- Weir, J. T. & Price, T. D. Limits to speciation inferred from times to secondary sympatry and ages of hybridizing species along a latitudinal gradient. *Am. Nat.* **177**, 462–469 (2011).
- Goldberg, E. E. & Lande, R. Ecological and reproductive character displacement on an environmental gradient. *Evolution* **60**, 1344–1357 (2006).
- Harvey, P. H. & Rambaut, A. Comparative analyses for adaptive radiations. *Phil. Trans. R. Soc. Lond. B* **355**, 1599–1605 (2000).
- Price, T. D. & Bouvier, M. M. The evolution of F1 postzygotic incompatibilities in birds. *Evolution* **56**, 2083–2089 (2002).
- Rosenzweig, M. L. *Species Diversity in Space and Time* (Cambridge Univ. Press, 1995).
- Endler, J. A. *Geographic Variation, Speciation, and Clines* (Princeton Univ. Press, 1977).
- Chesson, P. Mechanisms of maintenance of species diversity. *Annu. Rev. Ecol. Syst.* **31**, 343–366 (2000).
- Schluter, D. *The Ecology of Adaptive Radiation* (Oxford Univ. Press, 2000).
- Grant, P. R. & Grant, B. R. *How and Why Species Multiply* (Princeton Univ. Press, 2008).
- Yoder, J. B. *et al.* Ecological opportunity and the origin of adaptive radiations. *J. Evol. Biol.* **23**, 1581–1596 (2010).
- Johansson, U. S. *et al.* Build-up of the Himalayan avifauna through immigration: a biogeographical analysis of the *Phylloscopus* and *Seicercus* warblers. *Evolution* **61**, 324–333 (2007).
- Richman, A. D. & Price, T. Evolution of ecological differences in the Old World leaf warblers. *Nature* **355**, 817–821 (1992).
- Ghosh-Harihar, M. & Price, T. D. A test for community saturation along the Himalayan bird diversity gradient, based on within-species geographical variation. *J. Anim. Ecol.* **628**–638 (2014).
- Jönsson, K. A., Fabre, P. H., Ricklefs, R. E. & Fjeldså, J. A major global radiation of corvid birds originated in the proto-Papuan archipelago. *Proc. Natl Acad. Sci. USA* **108**, 2328–2333 (2011).
- Päckert, M. *et al.* Horizontal and elevational phylogeographic patterns of Himalayan and Southeast Asian forest passerines (Aves: Passeriformes). *J. Biogeogr.* **39**, 556–573 (2012).
- Kennedy, J. D. *et al.* Ecological limits on diversification of the Himalayan core Corvoidea. *Evolution* **66**, 2599–2613 (2012).
- Price, T. D. The roles of time and ecology in the continental radiation of the Old World leaf warblers (*Phylloscopus* and *Seicercus*). *Phil. Trans. R. Soc. Lond. B* **365**, 1749–1762 (2010).
- Wiens, J. J., Parra-Olea, G., Garcia-Paris, M. & Wake, D. B. Phylogenetic history underlies elevational biodiversity patterns in tropical salamanders. *Proc. R. Soc. Lond. B* **274**, 919–928 (2007).
- Price, T. D. *et al.* Determinants of northerly range limits along the Himalayan bird diversity gradient. *Am. Nat.* **178**, S97–S108 (2011).
- MacArthur, R. H. Patterns of species diversity. *Biol. Rev. Camb. Philos. Soc.* **40**, 510–533 (1965).
- Ricklefs, R. E. Species richness and morphological diversity of passerine birds. *Proc. Natl Acad. Sci. USA* **109**, 14482–14487 (2012).
- Rosenzweig, M. L. & Abramsky, Z. in *Species Diversity in Ecological Communities* (eds Ricklefs, R. E. & Schluter, D.) 52–65 (Univ. Chicago, 1993).
- Rabosky, D. L. & Matute, D. R. Macroevolutionary speciation rates are decoupled from the evolution of intrinsic reproductive isolation in *Drosophila* and birds. *Proc. Natl Acad. Sci. USA* **110**, 15354–15359 (2013).
- Cornell, H. V. Is regional species diversity bounded or unbounded? *Biol. Rev. Camb. Phil. Soc.* **88**, 140–165 (2013).
- Ricklefs, R. E. & Bermingham, E. The concept of the taxon cycle in biogeography. *Glob. Ecol. Biogeogr.* **11**, 353–361 (2002).
- Drummond, A. J. & Rambaut, A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**, 214 (2007).
- Harmon, L. J., Weir, J. T., Brock, C. D., Glor, R. E. & Challenger, W. GEIGER: investigating evolutionary radiations. *Bioinformatics* **24**, 129–131 (2008).
- Helmus, M. R., Bland, T. J., Williams, C. K. & Ives, A. R. Phylogenetic measures of biodiversity. *Am. Nat.* **169**, E68–E83 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank the government of India and the Chief Wildlife Wardens of the six Indian Himalayan states for permits. We also thank S. Dalvi, K. Jamdar, N. Jamdar, E. Scordato and D. Wheatcroft for help in the field, U. Borthakur and V. Mathur in the laboratory and E. Goldberg, R. Hudson, J. Kennedy, M. McPeck, A. Phillimore, D. Schluter, T. Tyrberg and J. Weir for advice. Tissue and toe pads for this project were provided by J. Cracraft (American Museum of Natural History), N. Rice (Academy of Natural Sciences, Philadelphia), M. Adams (The Natural History Museum, Tring), J. Dumbacher and M. Flannery (California Academy of Sciences), J. Bates and D. Willard (Field Museum of Natural History, Chicago), Herman Mays (Cincinnati Museum); M. Wink (Institut für Pharmazie und Molekulare Biotechnologie, Heidelberg); R. Brumfeld and D. Dittmann (Louisiana State Museum of Zoology); S. Edwards (Museum of Natural History, Harvard); G. Frisk (Swedish Museum of Natural History, Stockholm); J. Dean and J. Rappole (National Museum of Natural History Smithsonian); S. Birks (Burke Museum, University of Washington); the Zoologisches Forschungsmuseum Alexander Koenig, Bonn; J. Bolding Kristensen and J. Fjeldså (Zoological Museum, Copenhagen); the Zoologische Staatssammlung München, Munich; and K. Zyskowski (Yale University). This work was supported in part by grants from the US NSF and the National Geographic Society (TDP), the Jörnval Foundation, a Chinese Academy of Sciences Visiting Professorship (to P.A.), the Swedish Research Council (grants to U.O. and P.A.), the Wenner-Gren Foundation (a grant to U.S.J.), the Feldbausch Foundation of Mainz University (a grant to J.M.) and the German DFG (grants to B.H. and D.T.T., grant number Ti 679/1-1). We thank D. Tautz for making laboratory facilities available in Germany.

Author Contributions Study design and fieldwork: T.D.P. and D.M. Logistics: D.M. and P.S. Field collections: D.T.T., D.M.H., U.S.J., P.A., U.O., F.I., J.M., D.M. and T.D.P. Sequencing: D.M.H., U.S.J., D.T.T., U.O., P.A. and B.H. Arthropod censuses: M.G.-H. and T.D.P. Phylogeny construction: D.M.H., with early input from D.T.T. and P.A. Museum measurements: C.D.B., D.M., T.D.P. and U.S.J. Analysis and manuscript preparation: T.D.P., with input from B.H., D.M., D.M.H., D.T.T., P.A., U.S.J. and U.O.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to T.D.P. (pricedt@uchicago.edu).

METHODS

Conceptual basis and null models. This paper builds an estimate of the phylogenetic relationships of all members of a clade that are found in a locality, rather than the clade in total. Previous studies of adaptive radiations, defined here as divergence from a single ancestor to fill a variety of ecological niches³⁴, have studied clades, not communities, but these do not directly allow an assessment of how ecological differentiation—the essence of adaptive radiation—controls diversification and speciation.

The conceptual framework is illustrated in Extended Data Fig. 1, and follows refs 2, 3 and 35. Although adaptive radiation in young environments has been well studied, and connections between ecological diversification and reproductive isolation well established³⁶, little work has been done in older environments, which are a priori more likely to be ecologically full. As shown in Extended Data Fig. 1, the predictions are that speciation slows when ecological niches are filled, and the reason for this is that range expansions into sympatry become increasingly difficult. Range expansions may also slow because reproductive isolation is acquired more slowly when ecological opportunity is reduced (that is, different populations are now not under selection to adapt to different environments because niche space is increasingly filled up). However, even if populations are not subject to any divergent selection pressures at all, they should eventually differentiate to the level of full species. At that point, range expansions may remain difficult because of competition from the sister species (step 4 in Extended Data Fig. 1), as well as from other species in an environment where niches are increasingly occupied.

Our conclusions are that east Himalayan birds have approached this last step. If sympatric species were about the same age as the attainment of reproductive isolation, incomplete reproductive isolation may be limiting species accumulation. However, the average time to the common ancestor of pairs of close relatives (sisters on the assemblage tree) living in the sympatric assemblage is long (7 Myr). By contrast, postmating reproductive isolation in nature may typically be completed by 3 Myr since separation from their common ancestor (although complete loss of fertility in hybrids can evidently take longer than this⁶) because, first, many deleterious effects in hybrids are apparent by that time, which will lower fitness in the wild, and second, hybrids between species often fall between niches occupied by the parental forms^{2,10,11,36}. This reasoning applies to postmating isolation and it is possible that premating isolation remains incomplete. If so, complete loss of hybrid fitness plus ongoing hybridization would be a powerful means of setting range limits⁴. However, hybrid zones between species separated by more than 3 Myr are rare^{2,3}. More specifically, in our extensive population studies of thousands of breeding pairs of songbirds in the western Himalayas³⁷, we have never observed hybridization events between species separated by more than 4 Myr. Further, we noted in the main text that 90% of the species absent from the study area do overlap with their closest relative in the study area somewhere in the Himalayas, implying that premating reproductive isolation is essentially complete between them.

We bolster the argument of limits through niche-filling by building a time-dated phylogeny of all species in the community, and find both a substantial slowdown in per-lineage rates of splitting within the phylogeny and a slowdown in morphological evolution over time. Both of these results are consistent with niche availability limiting diversification, but they need to be evaluated against a suitable null model:

Lineage splitting. A slowdown in lineage splitting refers to a less than exponential increase in species, as expected from a pure birth (Yule) model (a model with constant speciation and extinction leads to a greater-than-exponential increase³⁸). A slowdown in lineage splitting has been observed in many studies of large clades³⁹ and has sometimes been interpreted as a consequence of niche-filling⁴⁰, as we do here. One bias is that there is a lag time to the production of allopatric forms that will then go on to be species, with the result that some of the recent splits in the phylogeny are not included when they should be^{39,41,42}. Previous work has addressed this difficulty by simply lopping off the most recent branch length (for example, refs 39 and 41). An advantage of working with sympatric assemblages is that species are defined unambiguously, so the lag time to species recognition is not an issue. However, given allopatric/parapatric speciation, there must be a delay to sympatry that includes the following features: (1) development of reproductive isolation between allopatric forms, (2) dispersal across any barriers, and (3) establishment in the community. As can be seen from the plot of lineage diversification versus time in Fig. 2, and as quantified in Extended Data Table 1, a strong slowdown remains even if the phylogeny is truncated 3 Myr ago, which is surely longer than an appropriate null model for how long a lag time should be (it is longer than it has taken to produce the entire Darwin's finch radiation², which contains up to 11 sympatric species).

Sampling. One set of null models considers that the Himalayan assemblage represents a random sample from some larger region in which the Himalayas are embedded. A random sample of species from the tips results in a greater slowdown than is observed in the larger phylogeny, because deep nodes are generally captured, but shallow nodes often will not be⁴³. Whether this is an appropriate null model is debatable

(see below) but we test it in this section. We consider sampling from both all Asian oscines and the pure birth model, and find the downturn in the study area to be much stronger than expected on the basis of random sampling from these larger species pools, even if the plot of lineage versus time is truncated at 3 Myr ago (Extended Data Table 1). In addition, trees sampled from a larger species pool (based on the tree of Jetz *et al.*³¹) have ages of sisters not greatly different from those in the larger pool (4.5 Myr compared to about 4 Myr in the analyses of Extended Data Table 1), which is substantially shorter than is observed in the sympatric assemblage (see the 'Phylogeny construction' section, in which the tree of ref. 31 is compared to the oscine tree illustrated in Fig. 1b).

Given that neither a 3-Myr lag time to sympatry nor sampling can account for the strong slowdowns, we argue that the rejection of a constant diversification rate in the phylogeny of sympatric species results because more recently produced allopatric species find it increasingly difficult to invade the community, plus possibly there has also been a slowdown in the global production of species, which, by induction, is itself ultimately driven by the failure of range expansions. Because over 1,500 songbird species are present in Asia but only 358 are present in the study area, a failure of establishment (range expansions) clearly accounts for some of the long lag times.

Alternative null models. Despite the analyses reported in Extended Data Table 1, we suggest that random sampling from a larger phylogeny is not an appropriate null model if the goal is to assess ecological controls on species diversification. This is because most related allopatric forms reflect either ecological replacements, or responses to ecological diversity in other locations of the world and therefore are not part of an assemblage responding to local ecological controls. We illustrate the reasoning by considering one species present in the Himalayas, *Zosterops palpebrosus*, together with its closest sequenced relatives, as taken from the tree of ref. 31 (Extended Data Fig. 2).

Most of the members of the *Z. palpebrosus* clade (the clade containing the upper ten species in Extended Data Fig. 2) consist of allopatric forms, some of whom are quite young and may not be reproductively isolated, and all of which occupy an arboreal fruit-, flower- and insect-eating niche, and commonly forage in flocks, that is, are ecological equivalents in different locations. We suggest that young allopatric forms reflect species turnover within the same ecological niche⁴⁴, with extinction attributable to both small population size and recent climatic change, followed by colonization and population fragmentation creating peripheral forms. These allopatric forms would not be recognized as different species in the fossil record⁴⁵ and depend on expansion out of more stable refuges, such as the eastern Himalayas is inferred to have been^{15,21}. This then suggests that a local-to-regional focus (stage 5 of Extended Data Fig. 1) as a limit on regional species numbers is at least as appropriate as the common regional-to-local focus as a limit on local diversity (compare ref. 46). The regional-to-local focus forms the basis for considering the local fauna to be a sample from the greater region, as tested in a previous section. However, if local interactions prevent the build-up of species in sympatry, the regional fauna is ultimately limited by local interactions, rather than the other way round.

Extended Data Fig. 2 also indicates some sympatry among *Zosterops* species. First, a pair of species on Reunion and Mauritius is thought to be the result of a double invasion⁴⁷ where one species (*Z. olivacea*) has become a flower specialist and is solitary⁴⁸. Second, a clade of allopatric forms (the lower five species in Extended Data Fig. 2) contains some species sympatric with *Zosterops palpebrosus*. In sympatry, these species differ ecologically from *Z. palpebrosus*: in China, *Z. japonicus* is migratory, and species in the Indonesian islands are altitudinally segregated from *Z. palpebrosus*. Thus, these examples of sympatry may be considered a response to ecological conditions in different regions. Assessment of ecological controls on diversification would need to include both the timing of colonization and ecological amplitude in these regions.

Models of ecological controls. In young environments, a large diversity of niches may appear more or less simultaneously, with subsequent rapid diversification to fill these niches, as in classic examples such as cichlid fish and Darwin's finches. In other examples, opportunity may arise in pulses, or uniformly (for example, as a result of extinctions). Price³² and Harvey and Rambaut⁵ considered a simple model in which new niches arose uniformly through time (for example, as a result of an occasional extinction) and species evolve to occupy the new niche from the ecologically closest ancestor. As shown in Extended Data Fig. 3, this model creates a strong downturn in the lineage splitting pattern, as well as a burst of ecological and morphological diversification early in the radiation³².

The results of the ecological controls model illustrated in Extended Data Fig. 3 are broadly consistent with the pattern we observe for lineage splitting and morphological evolution in the east Himalayan assemblage (Fig. 2). However, elevation does not show this pattern, with disparity accumulating relatively late. We suggest that the best interpretation is that a burst of ecological opportunity was relatively recently created along the elevation axis, in association with climate change 10 Myr

to 6 Myr ago¹⁹ (the mammalian fossil record of Pakistan shows a high turnover across this period⁴⁹).

Although morphological disparity in the 358 east Himalayan species phylogeny rapidly accumulates (Fig. 2, Extended Data Fig. 4), it still does so more slowly than the disparity curve of the simple niche-filling model illustrated in Extended Data Fig. 3. One contributing reason appears to be that the relatively recent appearance of high-elevation habitat creates opportunities along the other (body size and shape) axes, which are not always taken up by colonization from the ecologically most similar relative (an assumption of the model). A striking example is the high-elevation ground tit (*Pseudopodoces humilis*), which is the rightmost point in the tit clade of Fig. 2 (*Parus* and *Pseudopodoces*, shown in green). Far larger than other tits, *Pseudopodoces humilis* is morphologically convergent on low-elevation mynas: both *P. humilis* and mynas are predominantly open-country ground foragers. We speculate that climatic, habitat (absence of open areas at mid-elevations), and geographic barriers caused the 'open' niche at higher elevations to be not easily invaded by mynas, triggering the evolution of the ground tit from a tree-living *Parus*. This species is separated by an estimated 8 Myr from its closest relative in the study area, and the speciation occurred about the time high-elevation radiations were taking place in other groups 10–6 Myr ago (see Extended Data Fig. 6).

Significance tests of morphological evolution. For the main analysis in the text, we used the natural logarithm of mass— $\ln(\text{mass})$ —and the first two principal components (PC1, PC2) from the residuals of $\ln(\text{each morphological trait})$ on $\ln(\text{mass})$, using the entire Himalayan oscine data set for which we have measurements, $N = 441$ species. Correlations with the original variables are in Extended Data Fig. 5. We standardized each of these three variables— $\ln(\text{mass})$, PC1 and PC2—to have unit variance. The motivation for the standardization is that subtle differences in shape can have as much ecological significance as more obvious changes in size^{14,18} making it more meaningful to give each equal weighting in morphological dispersion assessments. However, similar results on morphological volumes accrue if the values are unstandardized, in which case size becomes dominant.

The time course of morphological evolution is illustrated in Extended Data Fig. 4, replicating Fig. 2, but with diversification expected under the Brownian-motion model added. We used two methods to assess the significance of the slowdown in morphological evolution. First, we fitted two-rate Brownian-motion models for morphological traits to the east Himalayan oscine phylogeny using Geiger (Extended Data Fig. 5a, b). The two-rate model includes two additional parameters over the one-rate model: the breakpoint and the rate difference before and after the breakpoint. To account for uncertainty in tree topology in assessment of significance, we averaged likelihood scores over 100 trees sampled from the Bayesian posterior distribution of trees, and to assess significance we used the likelihood profile (Extended Data Fig. 5b).

In a second approach, we compared Ornstein–Uhlenbeck models to Brownian motion models (Extended Data Fig. 5c). This analysis was modified in several ways from the analysis in the previous paragraph. First, unlike the disparity plots (Fig. 2 and Extended Data Fig. 4) and the assessment in Extended Data Fig. 5a, we extracted three phylogenetically corrected principal-component scores⁵⁰ from the correlation matrix of the log-transformed morphological measurements, for the eastern oscine species alone ($N = 355$ species for which we have measurements, out of 358 species in total). The correlations ("revealed PCs") with the original variables are in Extended Data Fig. 5d. We then asked at what point in time morphological evolution becomes constrained, by fitting Ornstein–Uhlenbeck and Brownian motion models to all clades subtended above a certain timeline, plus the one additional clade from the root to that timeline. In the Ornstein–Uhlenbeck models, all lineages subtending more than one species at each timeline were modelled with separate optima, but the constraining parameter was assumed to be the same for each clade (compare ref. 51). In all models the dispersive parameter was set the same throughout the tree. Relative weights of the two models were computed from the corrected (second-order) Akaike Information Criterion (AICc) scores obtained in the R program OUCH⁵². For both the Ornstein–Uhlenbeck and Brownian-motion models, at every timeline those lineages subtending just a single present-day species were removed before the analysis.

Both the tests indicate a strong slowdown in morphological evolution towards the present. This slowdown is not simply a result of a slowdown in lineage splitting, with morphological evolution concentrated at speciation events⁵³. Over half of all speciation events occur after the slowdown in morphology has largely happened. These speciation events are often associated with elevation splits, ecological differences that are accompanied by relatively little morphological differentiation. As noted above, the slowdowns in morphological evolution are consistent with early diversification to fill niches associated with the major modes of life, as expected if ecological controls form the ultimate limit on species diversity and speciation.

Comparing morphology and elevation. The mid-point of elevation has evolved more recently than body size and body shape (Fig. 2, Extended Data Fig. 4). This is shown also by phylogenetic signal in the data, as measured using Blomberg's

K value^{54,55}. K is greater than 1 for size and shape (1.74 for size and 1.74 for shape, that is, close relatives are more similar than expected under the model of Brownian motion), and less than 1 for elevation (0.63, that is, there is more evolution in the terminal branches than expected under Brownian motion). Phylogenetic signal, as assessed by comparing the variance of phylogenetic contrasts in a randomization test⁵⁵ is significant for size and shape ($P = 0.02$), but not for elevation ($P = 0.07$).

To compare patterns of evolution for any pair of traits (for example, body size with elevation) we adapted the method of ref. 56. We computed the absolute values of the phylogenetic contrasts throughout the tree, and regressed them on node height: a positive value indicates more evolution at the tips. Regression slopes are negative for size and shape and positive for elevation. We then compared the slopes of the regressions with two-sample Student's t -tests (with 351 degrees of freedom, as morphology was not available for four of the 358 species, implying 353 contrasts). The results were: size versus elevation, $t = 3.9$, two-tailed $P < 0.001$; size versus shape, $t = 1.34$, $P = 0.18$; and shape versus elevation, $t = 2.8$, $P = 0.006$. All traits were standardized to unit variance before the analysis, to remove scale effects.

Phylogeny construction. For taxonomic sampling, we incorporated all 461 oscines, all 5 suboscines and 133 of the 149 nonpasserines (from the Piciformes, Bucerotiformes, Upupiformes, Trogoniformes, Coraciiformes, Cuculiformes, Psittaciformes, Apodiformes, Strigiformes and Columbiformes, excluding Galliformes, Anseriformes and Falconiformes) found in the Himalayas^{21,57}. The resulting data set includes over 120 species sequenced for the first time, including rare species such as the Bugun liocichla (*Liocichla bugunorum*), known only from a few pairs. We also included 37 non-Himalayan species for the purpose of time-calibration (see source data for Fig. 1). Taxonomy follows ref. 57, except as outlined in the source data file associated with Fig. 1.

The majority of new sequences in this study were amplified using tissues sourced from museum collections or from feathers and blood collected in the field. However, we had to rely exclusively upon material from toe-pads for 54 species (31 passerines and 23 nonpasserines) without available fresh material. Owing to the degraded nature of DNA in these samples, amplification of mitochondrial loci proved difficult and nuclear data often impossible (see DNA extraction, amplification and sequencing). To alleviate this limitation and maximize the amount of data used for phylogenetic analyses, in some cases we incorporated genetic data from non-Himalayan surrogate species. We incorporated sequence data from surrogate species, as listed in the source data for Fig. 1, according to three different rationales: (1) if a species was the only member of its genus in the Himalayas we used an extra-limital congener; (2) if a species' genus contained more than one species in the Himalayas but the non-Himalayan sister species of that particular species was clear and (3) if a genus was not monotypic in the Himalayas and no locus-specific data for that genus was available for any of its Himalayan members, locus data from a non-Himalayan representative was assigned to a single Himalayan representative.

For example, the genus *Oriolus* contains four species that inhabit the Himalayas, for none of which we had adequate material to amplify the *RAG1* locus and none of which have sequence data for the *RAG1* locus on GenBank. Here, we added *RAG1* from the non-Himalayan *Oriolus cruentus* for the Himalayan *Oriolus kundoo* because we have no other *RAG1* data for this genus (rationale 3). Rationale 3 makes the assumption that the genus in question is monophyletic, which is most likely to be the case when further qualified by rationale 2. To assess whether the use of genetic data from surrogate species induced a bias in phylogenetic reconstruction, we repeated the phylogenetic analyses outlined below with all surrogate information removed. No bias was detected.

DNA extraction, amplification, and sequencing. DNA was extracted using Qiagen DNeasy Blood and Tissue kits (Qiagen) according to the manufacturer's instructions. Museum tissue and toe-pad materials were digested overnight in a mixture of 180 μl ATL buffer and 20 μl proteinase-K and then eluted with 200 μl or 50 μl of AE buffer respectively, following the manufacturer's instructions. We sequenced the mitochondrial cytochrome *b* gene (*cytb*), the mitochondrial NADH dehydrogenase subunit 2 (*ND2*), intron 2 of the myoglobin gene, introns 6 and 7 of the ornithine decarboxylase (*ODC*) gene, intron 11 of the glyceraldehyde-3-phosphodehydrogenase (*GAPDH*) gene, and the recombination activating protein 1 (*RAG1*) gene—partial coding sequences. The six loci were amplified and sequenced using standard primers and amplification profiles as described in ref. 58 for mitochondrial cytochrome *b*, by M. D. Sorenson *et al.* (<http://people.bu.edu/msoren/primers.html>) for mitochondrial *ND2*, ref. 59 for *myoglobin* and *GAPDH*, and ref. 60 for *RAG1*. We designed new primers to amplify the *ODC* locus for this study. Toe-pad samples were amplified as needed with a series of short, overlapping fragments of 200–600 base pairs, using a large set of genera-specific internal primers. All primers used in this study are available for use on request. Polymerase chain reaction (PCR) products were cleansed using ExoSap IT and sequenced directly on an ABI3730XL capillary sequencer.

Phylogenetic analyses. Sequences for each locus were aligned using MAFFT v7.1 (<http://mafft.cbrc.jp/alignment/software/>). Some manual adjustment was necessary

for nuclear sequences. Indels that were specific to a single species within the alignment were removed to reduce noise. The best-fitting model for each locus was identified with the AIC implemented in MrModelTest2 v2.3 (ref. 61), in conjunction with PAUP* (ref. 62), as follows: GTR+G for *cytb* and *myoglobin*; GTR+I+G for *ND2*, *ODC*, and *RAG1*; and HKY+G for *GAPDH*.

A time-calibrated phylogeny of the Himalayan Passeriformes was estimated by Bayesian inference using BEAST v1.8 (ref. 28) with computational assistance from BEAGLE v2.0 (ref. 63). For computational reasons, we ran the passerines (including the suboscines) largely independently from the remaining nonpasserines. Some species were included in both runs to help with time-calibration. Each locus was assigned its own partition with unlinked substitution and clock models but with a linked tree model. Thirteen time calibrations were used to date the tree. All dates were modelled with a uniform prior. Biogeographic calibrations were modelled between the date of geographic activity and the present to place a maximum age boundary. Fossil calibrations were modelled between 80.0 million years ago and the date of the fossil to place a minimum age boundary. By using multiple calibration points scattered broadly throughout the tree, we expect the accuracy of age calibration to improve as the average distance between calibrated and uncalibrated nodes decreases⁶⁴. We ran BEAST for 50 million generations sampling every 5,000 generations for a total of 10,000 trees. We assessed run performance and determined appropriate burn-in length using Tracer v1.5 (ref. 65) and constructed the maximum-likelihood clade credibility tree using TreeAnnotator v1.8 (ref. 28). Ten of the 13 time calibrations in the passerine part of the tree have been independently verified to be appropriate for use⁶⁶. We confirmed the additional three age calibrations using methods similar to those of ref. 66.

For the main phylogeny illustrated in Fig. 1b, a time-calibrated nonpasserine tree was run using the same BEAST methods, as described above. We constrained the order-level topology according to ref. 67 both because of the difficulty of achieving accurate order-level resolution with our limited data set and also because these deeper relationships are beyond the scope of this paper. We incorporated 11 time calibrations, three of which were also used for the oscine tree, using the same uniform prior logic for fossil and biogeographic calibrations. We spliced together the Passeriformes and nonpasserine trees using APE⁶⁸, based on the several passerine species included in the nonpasserine tree. Additional phylogeny appendices are available as follows: (1) passerine and nonpasserine trees submitted to <http://treebase.org> (<http://purl.org/phylo/treebase/phyloids/study/TB2:S15660>); (2) the XML file used to create the passerine tree in BEAST (Supplementary Information).

The Jetz *et al.*³¹ tree. The global tree of birds by ref. 31 includes all the species we studied here. However, among the 358 oscines in the eastern square, 85 species in their tree were without sequence data and were inserted into an assigned clade based on taxonomy, using the pure birth (Yule) model and nearly all the other species had less sequence data than we used. In our investigations we computed a maximum clade credibility tree on a sample of 100 trees ('Hackett backbone') downloaded from <http://birdtree.org>. The portion of the tree³¹ that includes the species we studied differs in several respects from ours and shows a much smaller slowdown in lineage-splitting events towards the present (the splitting-rate correlation coefficient $r = 0.13$ lineages per lineage per Myr over the first third of the tree to $r = 0.09$ to $r = 0.085$ over the last third; this compares with $r = 0.27$, $r = 0.13$ and $r = 0.06$ in the presented tree). The difference appears to be for at least three reasons: (1) the much older root of the oscines in the tree of ref. 31 (53 Myr ago versus 34 Myr ago in the current one), (2) the tree of ref. 31 inserted 25% of the species using the pure birth model because no sequence data at all was available and (3) taxonomy is a poor guide to where these species should be inserted.

The much older root in the tree of ref. 31 probably stems from the use of old fossils (in their analysis, young calibration dates could not be used, given the way clades were inserted into a backbone tree) and perhaps the use of log-normal priors, which gives undue weight to point estimates. The age for the root is incongruent with the fossil record, plus there is a relative lack of passerine morphological diversity⁶⁹. The root of the Eurasian species in our tree dates precisely to the Oligocene/Eocene boundary (34 Myr ago), a period of extreme cooling with a drop of several degrees Celsius in the oceans over a short time⁷⁰, and large turnover in the Asian mammalian fossil record⁷¹. It is therefore in accord with the idea that ecological opportunity drove songbird diversification.

Despite the conflicts between trees, a critical result on the age of 'sisters' (closest relatives on the phylogeny of the 358 sympatric oscine species) is consistent: the average age of sister pairs in the tree from ref. 31 (7 Myr \pm 0.47 Myr, $N = 112$ sisters, with 9 pairs separated by less than 2 Myr) is similar to that in the tree presented here (7.1 Myr \pm 0.3 Myr, $N = 116$, with 3 pairs separated by less than 2 Myr). The great age of Himalayan species appears to be robust.

Estimating diversity along the elevational gradient. For Fig. 3a we used the phylogenetic species diversity measure of Helmus and Ives^{70,72} because it captures phylogenetic structure in a way that is insensitive to phylogeny size. A value of 1 is equivalent to a star phylogeny, whereas a value close to zero indicates that many

species are close relatives. We pruned the larger phylogeny to a smaller one that connects all species thought to be in a particular elevational belt, based on overlapping altitudinal ranges (see source data for Fig. 2). Results were similar if we used species present in our censuses of 5-hectare plots instead.

In Extended Data Fig. 6 we show the plot of lineage versus time for the 500 m and 3,000 m elevation, both of which are estimated to contain the same number of species. The plot is cast on a linear scale to highlight the differences. The lower elevation has a more rapid accumulation of species from 20 Myr ago to 10 Myr ago, whereas the higher elevation has a more rapid accumulation from 10 Myr ago to 6 Myr ago. Note that although the higher elevation consists of younger species, very few species accumulate at either location over the past 6 Myr.

For Fig. 3b, as in the case of the phylogenetic diversity measure, we used all species thought to be in a particular elevational belt, based on overlapping altitudinal ranges. However, results were similar if we used species present in our censuses of 5-hectare plots instead (Extended Data Fig. 7). The volume of the convex hull was computed in morphological 3-space (Extended Data Fig. 5A, D) using the library FD^{73,74}. The convex hull is sensitive to sample size, hence is conservative with respect to the main finding (higher volume at higher altitudes, even though species numbers are less than at middle altitudes). A measure insensitive to sample size (summed variance of the three dimensions) gave congruent results (not shown). In Extended Data Fig. 7 we show morphological space on the first two dimensions at four locations. At highest elevations, both the shape and size dimensions are greatest, which may partly reflect an absence of nonpasserine competitors¹⁸ as well as more open habitat, which favours a relatively long tarsus.

Field methods. The 'study area' is bounded by the coordinates 88° E, 93.1° E, 26.3° N and 28.1° N, covers an area of 10,389 km², and spans an elevational range from 37 m to 6,778 m. Essential climatic features are illustrated in Extended Data Fig. 8. Temperature, precipitation and estimated productivity decline monotonically with elevation (Extended Data Fig. 8; ref. 75). Precipitation shows the greatest spatial variation at low elevations (Extended Data Fig. 8).

T.D.P. and D.M. collected data over an elevational range of 150–4,100 m in six field seasons (2007–2012, mostly in the later years), preceded by three (T.D.P.) or more (D.M.) earlier visits to familiarize ourselves with the birds. We worked entirely in India, with no visits to Bhutan, and set up 5-hectare grids using the Global Positioning System (GPS) in as undisturbed forest as possible, usually over 2 km from any substantial clearing. Locations of the grids are plotted in Extended Data Fig. 8, as listed in the Source Data for Fig. 4. Two observers walked slowly over a grid for two mornings during the breeding season, from dawn for about 4 h, mapping singing males and recording sightings. For the more common species, concordance among observers in the number of singing males was high. Although only one observer often recorded one or other of the uncommon species, both observers recorded similar numbers of species in total. We assume that a single observer's record approximates breeding birds on the grid (that is, rare species which were seen by one or other of us are often at lower density than 1 pair per 5 hectares and some are unlikely to be breeding on the grid.) We analysed the surveys of T.D.P. and the surveys of D.M. plus assistants separately and got similar results. T.D.P.'s surveys are reported (see source data for Fig. 4 for the listing of surveys included). M.G.-H. collected arthropod data on or near nine of the grids, during three separate visits^{76,77}. Methods usually involved placing fifty 200-litre garbage bags over lower branches, breaking the branch, anesthetizing the arthropods and then sorting through the bag (see ref. 77 for standard errors.) Because not all grids were surveyed in this way, we used a quadratic curve fit to predict numbers of arthropods at a given elevation. The relatively few arthropods at low elevations were unexpected, so in May 2013, we made additional collections at a new site, plus some additional collections at site B1. See the source data for details on these collections. Vegetation data were collected on each grid²¹ and are also reported in the source data for Fig. 4. The smoothed lines in Fig. 4 were fitted using the lowess procedure in R (<http://stat.ethz.ch/R-manual/R-patched/library/stats/html/lowess.html>).

Mid-elevation peak in species richness. Species richness at the 2-km² scale was estimated using the same raster as the <http://worldclim.org> 2.5-arcminutes resolution, which resulted in 5,246 squares across the study area. The points represent the total number of breeding species estimated for each square. To assign a species to a square, we derived its east and west range limits along the Himalayas from the maps of Rasmussen and Anderton⁵⁷ (as described in ref. 21), and its altitudinal limits from our own work (source data for Fig. 2). Smooth lines in Fig. 1 are General Additive Model (gam) fits using the mgcv library in R (<http://cran.r-project.org/web/packages/mgcv/index.html>). Critical additional packages used to construct that figure were Sp^{78,79}, Raster⁸⁰ and PBSmapping⁸¹.

Rahbek⁸² drew attention to the presence of a peak in bird species richness at mid-elevations, rather than a monotonic decline in species numbers with elevation. In a survey of 78 bird studies McCain⁸³ found that about half showed a mid-elevation peak. Several ecological explanations have been presented. Rahbek⁸² noted that mid-elevations in the neotropics have increased structural diversity, notably epiphytes,

McCain⁸³ found that species numbers often correlate with productivity (mountains that are dry at the base are those that tend to have mid-elevation peaks), Rosenzweig and Abramsky²⁴ noted more generally that some mid-elevation peaks may arise even if productivity declines, owing to competition with other groups, and Terborgh⁸⁴ showed that different guilds have different elevational patterns, which he related to different resource distributions. Our interpretation of the Himalayan gradient is consistent with available resources being the ultimate driver of the peak. The mid-elevation peak is restricted to insectivorous oscines, which is related to arthropod abundance: this in turn may result from both climatic factors (for example, a spring flush of leaf) and the presence of efficient competitors at lower elevations. However, several other hypotheses for the mid-elevation peak have been proposed⁸⁵. First, larger areas may accumulate more species⁸². Within the study area (Extended Data Fig. 8), the area within 500-m bands (between elevations of 200 m and 3,700 m) is roughly similar, albeit with a slight mid-elevation peak (Extended Data Fig. 9a), but overall the low elevations (Indian plains) and the high elevations (Tibetan plateau) have much more area than the middle elevations. Three other hypotheses, not described in the main text, are considered in more detail.

The first is a concern with data. If estimates of elevational ranges are simply based on connecting the highest and lowest records of a species (so-called range-through methods), then as one approaches the domain boundaries sampling error inevitably means that some species are missed⁸⁶. Our 5-hectare plot surveys are unaffected by this issue and these local estimates match the integrated elevational distributions (compare Figs 1 and 4). Further, elevational ranges are unlikely to be underestimates of the true ranges because (1) they integrate over multiple years, publications and elevational gradients and (2) isolated breeding birds beyond their normal range limits may be especially recorded in the literature. Reason (2) means that it is conceivable that our ranges are over-estimated, but we have done our best to eliminate this on the basis of our field experiences.

The second concern is that of a suitable null model against which to assess the mid-elevation peak. In one much-discussed model, whose null character has been debated^{87,88}, shuffling ranges of fixed sizes on a bounded domain inevitably leads to more species at mid-elevations⁸⁹. Some other null models give weak or no mid-elevation peaks⁸⁶. In the original null model a prominent peak at mid-elevations is most expected if elevational ranges are a medium to large fraction of the total domain⁸³. In our study, oscine elevational ranges average 1,154 m, which is about a quarter of the distance from the plains to the treeline (just 32 of the 96 oscines at 700 m are also found at 1,900 m and of the 155 at 1,900 m only 30 are also at 3,100 m; none of the species at 700 m are also at 3,100 m). The original mid-domain null model results in zero species at the boundaries⁹⁰. We simulated this version, sampling species range mid-points on the domain size most favourable for the null hypothesis (200 m, 3,800 m) and obtaining mean R^2 associations of real and simulated data of 0.66, when richness was evaluated at every 100 m. Regions at the extremes lay above the 95% confidence limits from the simulated data and the whole transect between 1,000 m and 3,000 m lay below the null model confidence limits; a relatively poor fit. Because species distributions spread into south India and a long way above the treeline it is unclear what domain size to use: tests on larger domain sizes gave lower R^2 values. The mid-domain null has been generally rejected across bird studies⁸³, including an application to the bird diversity gradient in the eastern Himalayas⁷⁵ and is clearly not compatible with the distributions of nonpasserines and those songbirds that are not insectivores, both of which show monotonic declines from low elevations (Figs 1 and 4).

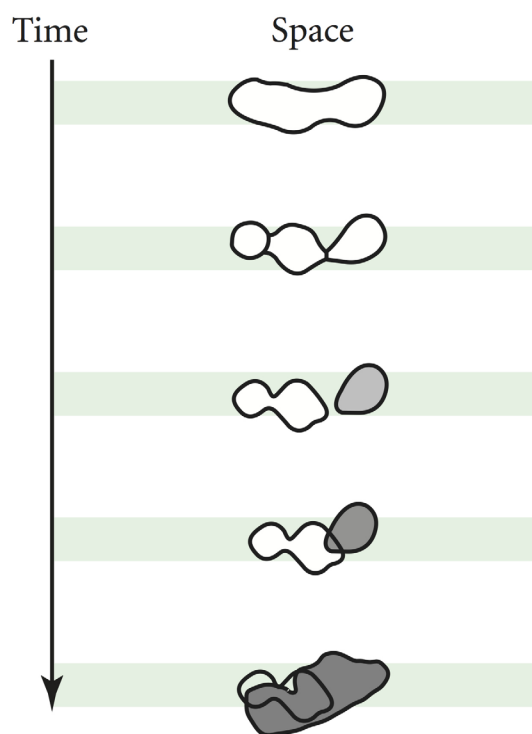
An important issue is whether the mid-elevation peak represents a “mass effect” whereby some species in a particular location are only present because of dispersal⁹¹, that is, they represent sink populations, which could not be maintained in the absence of immigration. In theory, species can disperse from both above and below at intermediate elevations but only from one direction closer to the domain edges. This may lead to a mid-elevation peak, with local plots saturated to similar levels but those at mid-elevations having excess sink species, that is, sink species⁹² occur to a greater extent in the mid-elevations. To address this, we focused on the local (grid) surveys, which show a mid-elevation peak comparable to that at the larger scales (Fig. 4a). We listed as plausible candidates for sink species all species represented by just a single breeding pair and which were under 300 m from either their upper or their lower elevational range limit (except in the case of grids below 300 m, when we used the criteria that these species should not extend south of the Himalayas into the plains or Assam hills). By this measure, up to 12 species at mid-elevations may be sink species. However, at elevations below 200 m up to 8 species may be sink species. If we remove sink species, the mid-elevation peak remains prominent (Extended Data Fig. 9b).

The mid-elevation peak is attributed mostly to a large number of small species (<15 g). This pattern is consistent with a large number of small arthropods at these elevations⁷⁷. At lower elevations, arthropod distributions are skewed towards larger sizes, and insectivores are also generally larger. Our hypothesis, following ref. 24, is that a large abundance of arthropods permits finer subdivision of the niche.

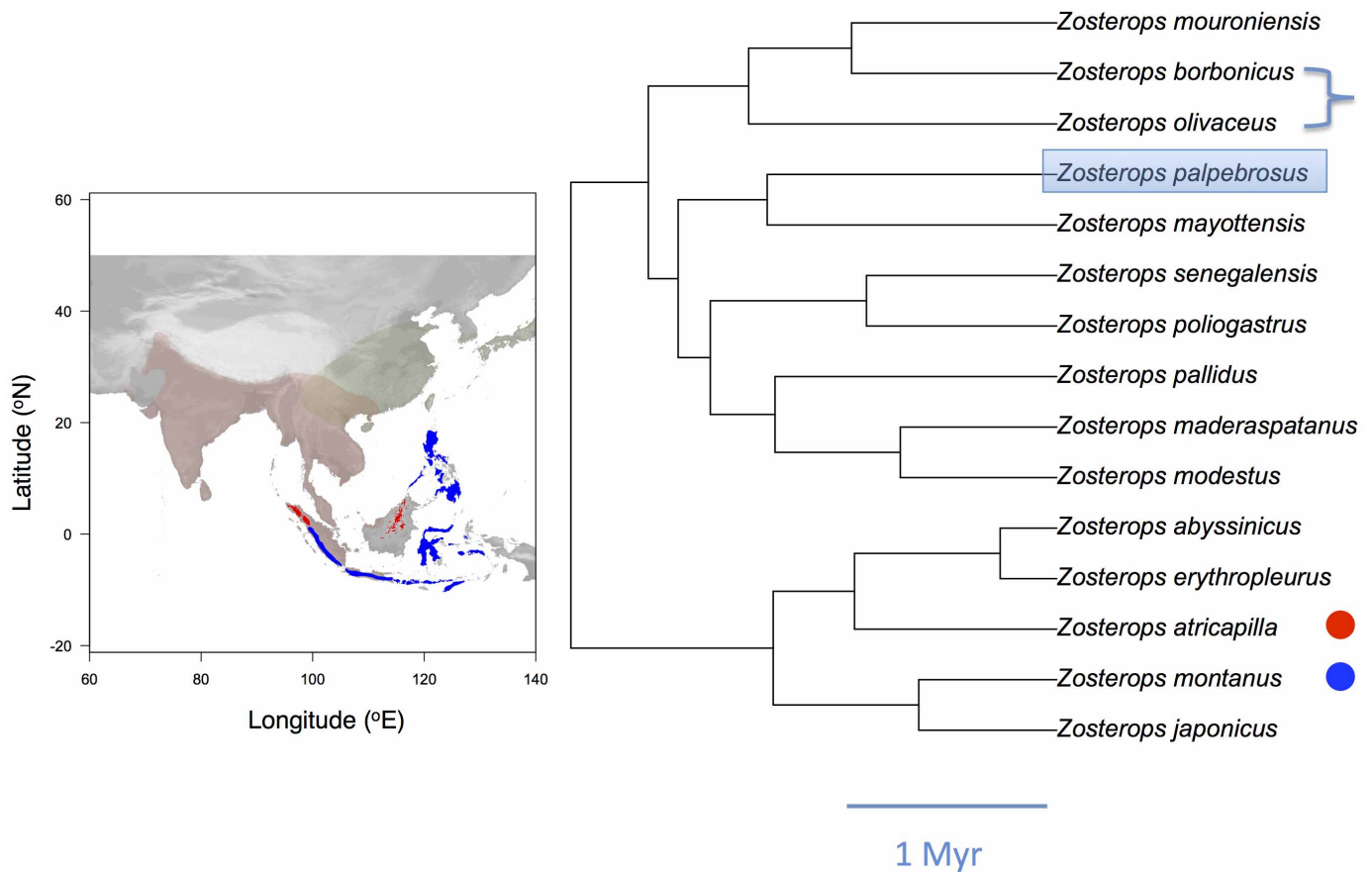
Terborgh and Faaborg⁹³ noted that insectivores may more generally be able to achieve fine niche partitioning, and this may well not be reflected in large morphological differences.

31. Jetz, W., Thomas, G. H., Joy, J. B., Hartmann, K. & Mooers, A. O. The global diversity of birds in space and time. *Nature* **491**, 444–448 (2012).
32. Price, T. Correlated evolution and independent contrasts. *Phil. Trans. R. Soc. Lond. B* **352**, 519–529 (1997).
33. Hijmans, R. J., Cameron, S. E., Parra, J. L., Jones, P. G. & Jarvis, A. Very high resolution interpolated climate surfaces for global land areas. *Int. J. Climatol.* **25**, 1965–1978 (2005).
34. Losos, J. B. Adaptive radiation, ecological opportunity, and evolutionary determinism. *Am. Nat.* **175**, 623–639 (2010).
35. Mayr, E. *Systematics and the Origin of Species from the Viewpoint of a Zoologist* (Columbia Univ. Press, 1942).
36. Nosil, P. *Ecological Speciation* (Oxford Univ. Press, 2012).
37. Price, T., Zee, J., Jamdar, K. & Jamdar, N. Bird species diversity along the Himalayas: a comparison of Himachal Pradesh with Kashmir. *J. Bombay Nat. Hist. Soc.* **100**, 394–409 (2003).
38. Harvey, P. H., May, R. M. & Nee, S. Phylogenies without fossils. *Evolution* **48**, 523–529 (1994).
39. Phillimore, A. B. & Price, T. D. Density-dependent cladogenesis in birds. *PLoS Biol.* **6**, 483–489 (2008).
40. Phillimore, A. B. & Price, T. D. in *Speciation and Patterns of Diversity* (eds Butlin, R. K., Bridle, J. & Schluter, D.) 240–256 (Cambridge Univ. Press, 2009).
41. Weir, J. T. Divergent patterns of species accumulation in lowland and highland neotropical birds. *Evolution* **60**, 842–855 (2006).
42. Etienne, R. S. & Rosindell, J. Prolonging the past counteracts the pull of the present: protracted speciation can explain observed slowdowns in diversification. *Syst. Biol.* **61**, 204–213 (2012).
43. Pybus, O. G. & Harvey, P. H. Testing macro-evolutionary models using incomplete molecular phylogenies. *Proc. R. Soc. Lond. B* **267**, 2267–2272 (2000).
44. McPeck, M. A. The ecological dynamics of clade diversification and community assembly. *Am. Nat.* **172**, E270–E284 (2008).
45. Rabosky, D. L. Diversity-dependence, ecological speciation, and the role of competition in macroevolution. *Annu. Rev. Ecol. Evol. Syst.* **44**, 481–502 (2013).
46. Ricklefs, R. E. Community diversity: relative roles of local and regional processes. *Science* **235**, 167–171 (1987).
47. Coyne, J. A. & Price, T. D. Little evidence for sympatric speciation in island birds. *Evolution* **54**, 2166–2171 (2000).
48. Gill, F. B. Ecology and evolution of sympatric Mascarene white-eyes, *Zosterops borbonica* and *Zosterops olivacea*. *Auk* **88**, 35–60 (1971).
49. Badgley, C. et al. Ecological changes in Miocene mammalian record show impact of prolonged climatic forcing. *Proc. Natl Acad. Sci. USA* **105**, 12145–12149 (2008).
50. Revell, L. J. Size-correction and principal components for interspecific comparative studies. *Evolution* **63**, 3258–3268 (2009).
51. Beaulieu, J. M., Jhwueng, D. C., Boettiger, C. & O’Meara, B. C. Modeling stabilizing selection: expanding the Ornstein-Uhlenbeck model of adaptive evolution. *Evolution* **66**, 2369–2383 (2012).
52. Butler, M. A. & King, A. A. Phylogenetic comparative analysis: a modeling approach for adaptive evolution. *Am. Nat.* **164**, 683–695 (2004).
53. Ricklefs, R. E. Cladogenesis and morphological diversification in passerine birds. *Nature* **430**, 338–341 (2004).
54. Blomberg, S. P., Garland, T. & Ives, A. R. Testing for phylogenetic signal in comparative data: behavioral traits are more labile. *Evolution* **57**, 717–745 (2003).
55. Kembel, S. W. et al. Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**, 1463–1464 (2010).
56. Mahler, D. L., Revell, L. J., Glor, R. E. & Losos, J. B. Ecological opportunity and the rate of morphological evolution in the diversification of Greater Antillean Anoles. *Evolution* **64**, 2731–2745 (2010).
57. Rasmussen, P. C. & Anderton, J. C. *Birds of South Asia: the Ripley Guide* (Lynx, 2005).
58. Weir, J. T. & Schluter, D. The latitudinal gradient in recent speciation and extinction rates of birds and mammals. *Science* **315**, 1574–1576 (2007).
59. Dong, F., Li, S. H. & Yang, X. J. Molecular systematics and diversification of the Asian scimitar babblers (Timaliidae, Aves) based on mitochondrial and nuclear DNA sequences. *Mol. Phylogenet. Evol.* **57**, 1268–1275 (2010).
60. Groth, J. G. & Barrowclough, G. F. Basal divergences in birds and the phylogenetic utility of the nuclear RAG-1 gene. *Mol. Phylogenet. Evol.* **12**, 115–123 (1999).
61. Nylander, J. A. A., Ronquist, F., Huelsenbeck, J. P. & Nieves-Aldrey, J. L. Bayesian phylogenetic analysis of combined data. *Syst. Biol.* **53**, 47–67 (2004).
62. Swofford, D. L. *PAUP*. Phylogenetic Analysis Using Parsimony (*and other methods)* Version 4 (Sinauer, 2003).
63. Ayres, D. L. et al. BEAGLE: an application programming interface and high-performance computing library for statistical phylogenetics. *Syst. Biol.* **61**, 170–173 (2012).
64. Linder, H. P., Hardy, C. R. & Rutschmann, F. Taxon sampling effects in molecular clock dating: an example from the African Restionaceae. *Mol. Phylogenet. Evol.* **35**, 569–582 (2005).
65. Rambaut, A. & Drummond, A. J. *Tracer v1.4* <http://beast.bio.ed.ac.uk/> (2007).
66. Weir, J. T. & Schluter, D. Calibrating the avian molecular clock. *Mol. Ecol.* **17**, 2321–2328 (2008).
67. Hackett, S. J. et al. A phylogenomic study of birds reveals their evolutionary history. *Science* **320**, 1763–1768 (2008).
68. Paradis, E., Claude, J. & Strimmer, K. APE: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**, 289–290 (2004).

69. Mayr, G. The age of the crown group of passerine birds and its evolutionary significance—molecular calibrations versus the fossil record. *Syst. Biodivers.* **11**, 7–13 (2013).
70. Hansen, J. *et al.* Target atmospheric CO₂: where should humanity aim? *Open Atmos. Sci. J.* **2**, 217–231 (2008).
71. Kraatz, B. P. & Geisler, J. H. Eocene-Oligocene transition in Central Asia and its effects on mammalian evolution. *Geology* **38**, 111–114 (2010).
72. Helmus, M. R. & Ives, A. R. Phylogenetic diversity-area curves. *Ecology* **93**, S31–S43 (2012).
73. Laliberté, E. & Legendre, P. A distance-based framework for measuring functional diversity from multiple traits. *Ecology* **91**, 299–305 (2010).
74. Laliberté, E. & Shipley, B. FD: measuring functional diversity from multiple traits, and other tools for functional ecology. *R package* version 1.0–11, <http://cran.r-project.org/web/packages/FD/index.html> (2011).
75. Acharya, B. K., Sanders, N. J., Vijayan, L. & Chettri, B. Elevational gradients in bird diversity in the eastern Himalaya: an evaluation of distribution patterns and their underlying mechanisms. *PLoS ONE* **6**, e29097 (2011).
76. Gross, S. & Price, T. Determinants of the northern and southern range limits of a warbler. *J. Biogeogr.* **27**, 869–878 (2000).
77. Ghosh-Harihar, M. Distribution and abundance of foliage-arthropods across elevational gradients in the east and west Himalayas. *Ecol. Res.* **28**, 125–130 (2013).
78. Pebesma, E. J. & Bivand, R. S. Classes and methods for spatial data in R. *R News* **5**, <http://cran.r-project.org/doc/Rnews/> (2005).
79. Bivand, R. S., Pebesma, E. J. & Gomez-Rubio, V. *Applied Spatial Data Analysis with R*. (Springer, 2008).
80. Hijmans, R. J. & van Etten, J. Raster: geographic analysis and modeling with raster data. *R package* version 1.9–92, <http://CRAN.R-project.org> (2011).
81. Schnute, J. T., Boers, N. M. & Haigh, R. PBS Mapping 2: user's guide. *Can. Tech. Rep. Fish. Aquat. Sci.* **2549**, 126 (2004).
82. Rahbek, C. The elevational gradient of species richness—a uniform pattern. *Ecography* **18**, 200–205 (1995).
83. McCain, C. M. Global analysis of bird elevational diversity. *Glob. Ecol. Biogeogr.* **18**, 346–360 (2009).
84. Terborgh, J. Bird species diversity on an Andean elevational gradient. *Ecology* **58**, 1007–1019 (1977).
85. Sanders, N. J. & Rahbek, C. The patterns and causes of elevational diversity gradients. *Ecography* **35**, 1–3 (2012).
86. Grytnes, J. A. & Vetaas, O. R. Species richness and altitude: a comparison between null models and interpolated plant species richness along the Himalayan altitudinal gradient, Nepal. *Am. Nat.* **159**, 294–304 (2002).
87. Hawkins, B. A., Diniz, J. A. F. & Weis, A. E. The mid-domain effect and diversity gradients: is there anything to learn? *Am. Nat.* **166**, E140–E143 (2005).
88. Zapata, F. A., Gaston, K. J. & Chown, S. L. The mid-domain effect revisited. *Am. Nat.* **166**, E144–E148 (2005).
89. Colwell, R. K., Rahbek, C. & Gotelli, N. J. The mid-domain effect: there's a baby in the bathwater. *Am. Nat.* **166**, E149–E154 (2005).
90. McCain, C. M. The mid-domain effect applied to elevational gradients: species richness of small mammals in Costa Rica. *J. Biogeogr.* **31**, 19–31 (2004).
91. Grytnes, J. A., Heegaard, E. & Romdal, T. S. Can the mass effect explain the mid-altitudinal peak in vascular plant species richness? *Basic Appl. Ecol.* **9**, 373–382 (2008).
92. White, E. P. & Hurlbert, A. H. The combined influence of the local environment and regional enrichment on bird species richness. *Am. Nat.* **175**, E35–E43 (2010).
93. Terborgh, J. W. & Faaborg, J. Saturation of bird communities in the West Indies. *Am. Nat.* **116**, 178–195 (1980).

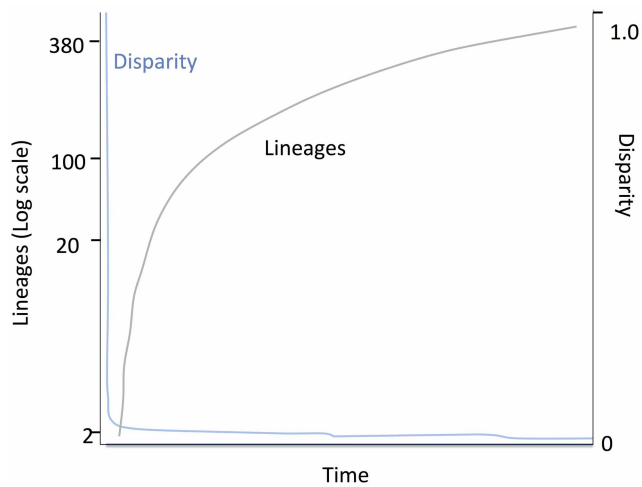


Extended Data Figure 1 | The speciation cycle. A species distributed across space becomes fragmented as a result of either vicariance (illustrated) or dispersal. After barrier formation, reproductive isolation develops. For the cycle to continue at least one of the species must expand into the other's range, which requires reproductive isolation, and generally also ecological compatibility.

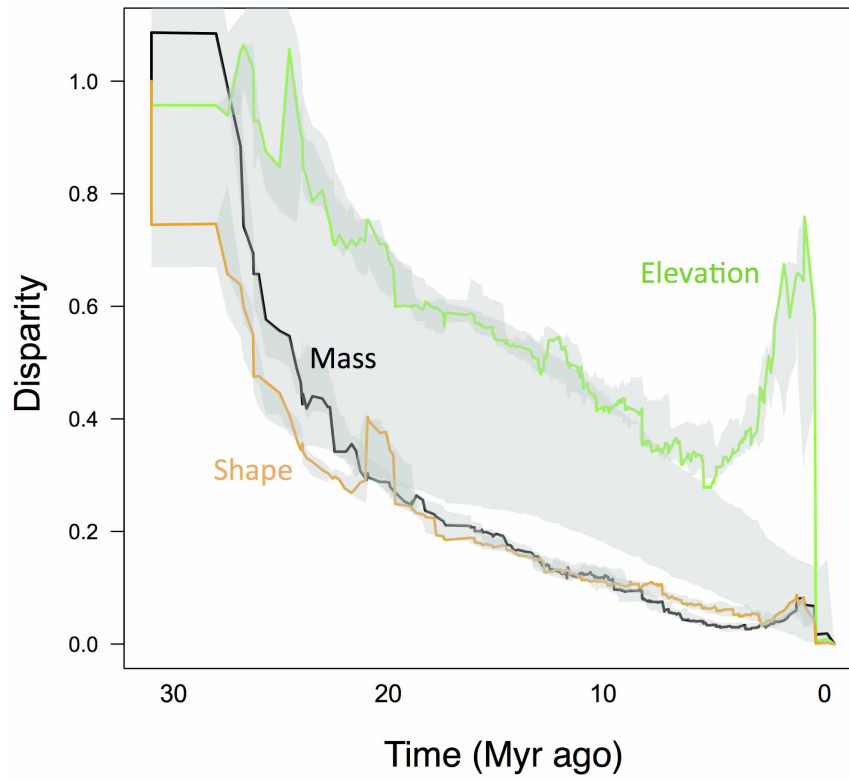


Extended Data Figure 2 | Close relatives of a single Himalayan species, the Oriental white-eye, *Zosterops palpebrosus*. Only species for which sequence data are available are included (tree from ref. 31). The range of *Z. palpebrosus* (light red) overlaps with members of the clade containing the lower 5 species, for example, with *Z. japonicus* (which is migratory)

(light green) in eastern China and *Z. atricapilla* (red) and *Z. montanus* (blue) in Indonesia, where *Z. palpebrosus* is altitudinally segregated from them. Within the *Z. palpebrosus* clade, all species are allopatric replacements, except for the two species on the Mascarene Islands (bracketed). The timeline is from ref. 31.

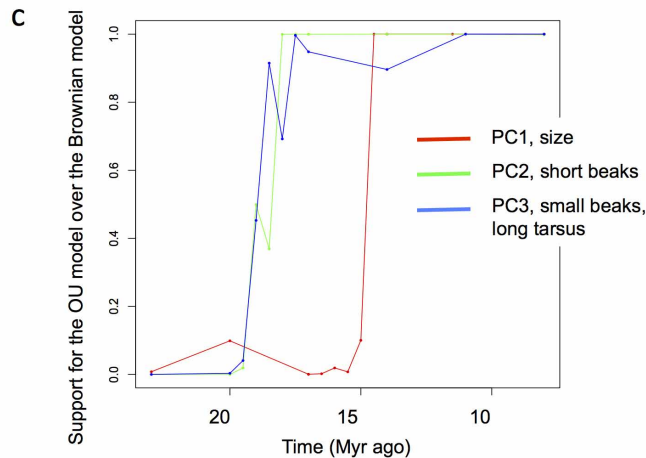


Extended Data Figure 3 | Plot of lineage versus time and morphological disparity generated in a simple model of ecological controls³². In this model, new niches appear uniformly through time, and new species arise to fill them, with the criterion that new species are always derived from the ecologically most similar form^{5,32}. For this simulation, the position of a new niche was drawn from a bivariate normal (x, y) with a correlation of 0.5, with 380 niches appearing sequentially and uniformly spaced in time. The result is a linear accumulation of species through time (that is, a downturn on the log scale), and most of the morphological variation accumulating early in the radiation (in the plot, disparity for one variable is shown).



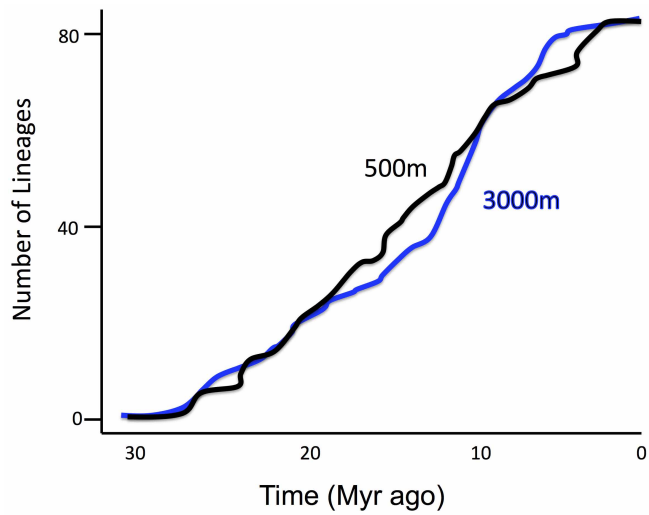
Extended Data Figure 4 | Disparity plots for morphology and habitat, with the null Brownian-motion model added. The large shaded area represents the 95% confidence limit from 100 simulations on 100 trees drawn from the

posterior distribution of the Bayesian analysis. The shaded areas around the data plots gives the 95% confidence limits based on phylogenetic uncertainty (based on the same 100 trees as above).

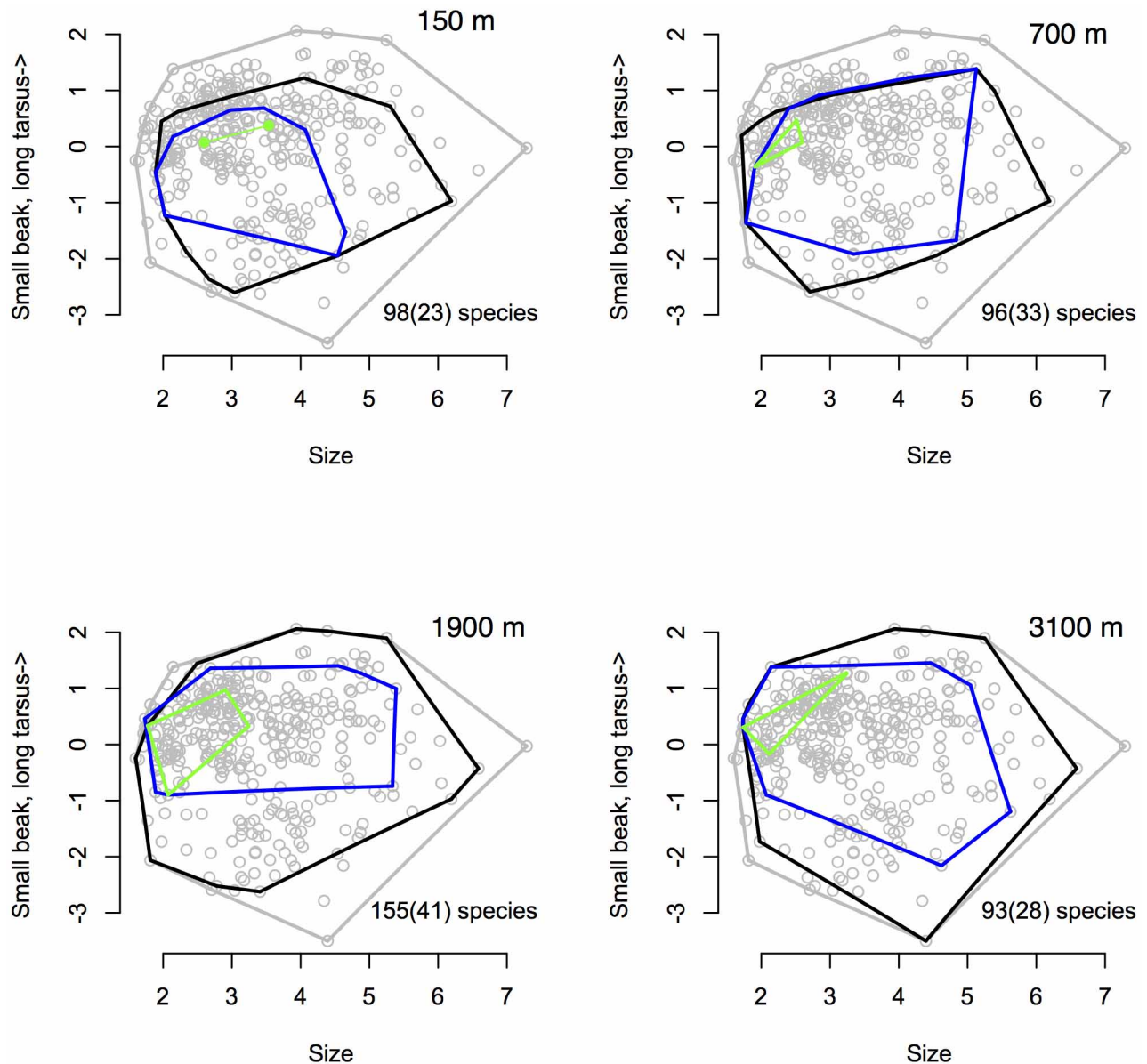


Extended Data Figure 5 | Slowdowns in morphological evolution across the tree of the east Himalayan oscines. **a**, Maximum likelihood breakpoints (the point in time at which one rate becomes favoured over the other) and changes in rate for two-rate models of morphological evolution. Significance values (* $P < 0.01$, ** $P < 0.001$) refer to likelihood ratio tests comparing the one- and two-rate Brownian motion models (PC2, $P = 0.16$). 95% support limits (parentheses) were derived from likelihood profiles averaged across 100 trees sampled from the posterior distribution of Bayesian trees. **b**, The likelihood profile for evolution of the first shape index (PC1). The likelihood for

each (x, y) combination was obtained as the average across 100 trees, then log-transformed for the figure. Numbers are the difference in log-likelihood from the maximum ($\times 100$). Only values less than 2 units are shown. The profiles are indicated by symbols (squares for the breakpoint, and circles for the rate difference). **c**, Relative weights of Ornstein–Uhlenbeck (OU) and Brownian-motion models of morphological evolution at different timepoints, based on phylogenetically corrected principal components (revellePCs) (see text). **d**, Correlations of PC scores with the original (log-transformed) variables.

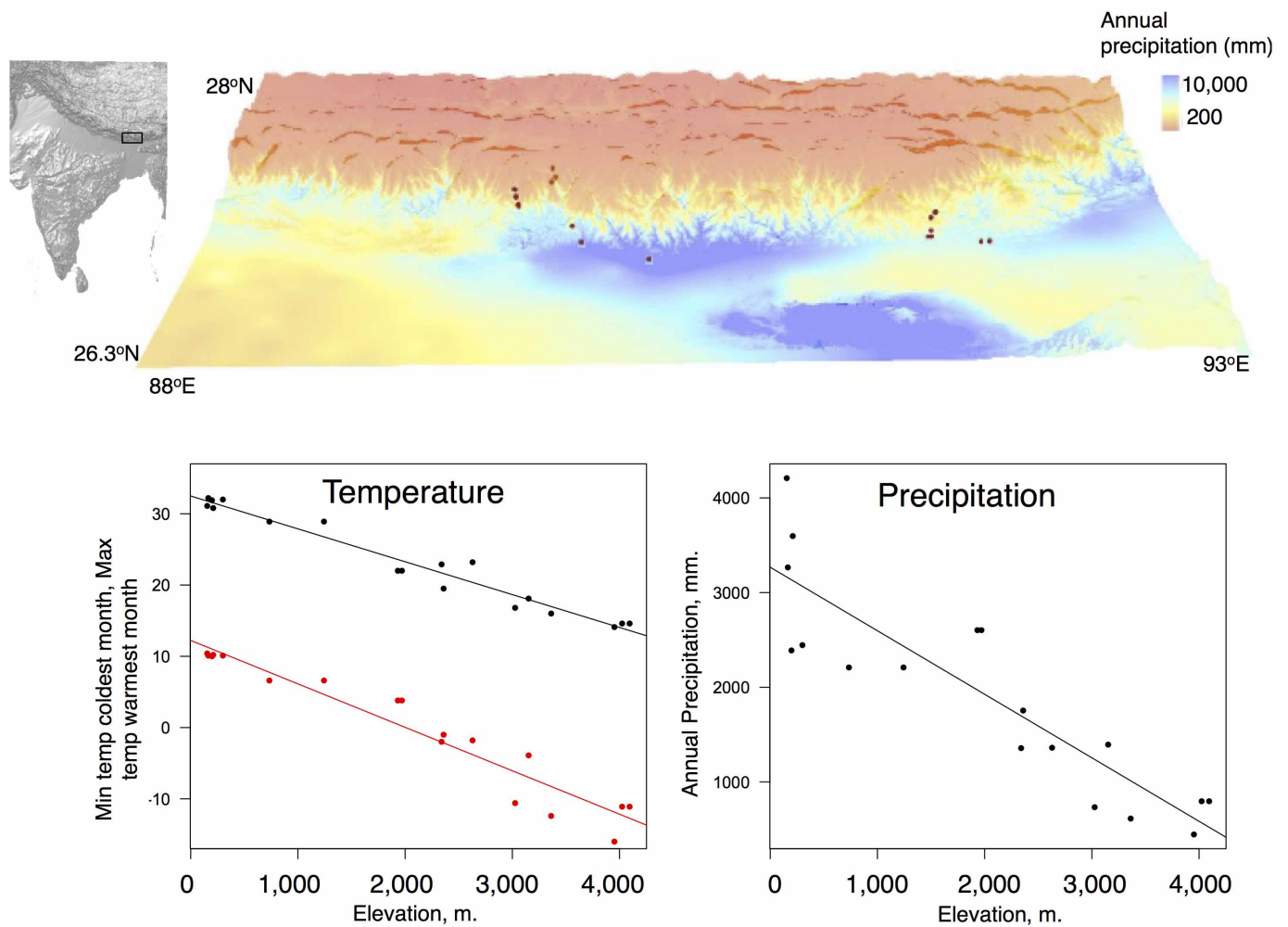


Extended Data Figure 6 | Plot of lineage diversity (on a linear scale) versus time for a phylogeny connecting all species present at 500 m and at 3,000 m. Eighty-two species are estimated to straddle each of these elevational bands.



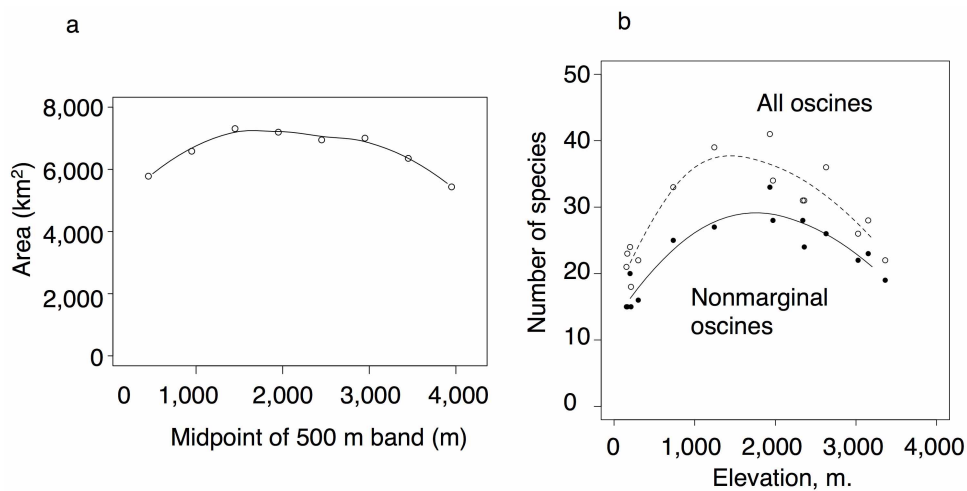
Extended Data Figure 7 | Morphology at specified elevations. Grey lines are the convex hull for all species in the study area (points as in Fig. 2). Black lines are the convex hulls for all species whose elevational ranges include the specified band. Blue lines are the convex hulls for all the species censused on 5-hectare grids at those elevations (see the source data, in order of elevation,

B2, A3, B1 and G1), and green lines are the convex hulls for all common (>5 pairs per hectare) species on those grids. Number of species is the number of all bird species in that elevational belt, plus (in parentheses) the number of songbirds censused on the grid.



Extended Data Figure 8 | Climate data (from <http://worldclim.org33>). The top panel shows precipitation mapped on to a topographical map of the study area, showing the locations of the 18 grids. The bottom panels show three

predicted climate variables (minimum and maximum temperatures, and precipitation) for the 18 grids. Lines are least-squares regression slopes.



Extended Data Figure 9 | Mid-elevation peak plots. The left panel shows area in 500-m bands between 200 m and 3,700 m in the study area (computed using <http://worldclim.org> altitude data). The right panel shows the number of

oscines in the censused 5-hectare grids and number of oscines in those grids discounted by possible sink species (rare species at the edge of their range).

Extended Data Table 1 | Significance in the downturn in the plot of lineage diversification versus time

Sample space	One tailed P (γ)
*1,504 Asian oscines	$P < 0.001$ (−2.80)
*1,662 Asian oscines	$P < 0.001$ (−2.61)
871 tips pure birth model	$P < 0.001$ (−2.95)
1,546 tips pure birth model	$P < 0.001$ (−3.12)
7,686 tips pure birth model	$P < 0.001$ (−3.49)

The significance was assessed by the γ statistic, where a value of less than −1.96 is considered a significant slowdown, when compared to the pure birth (Yule) model^{39,43}. The P value was assessed by drawing 1,000 random samples of 358 species from the indicated larger phylogenies; the average γ value from the simulations is in parentheses. The asterisk indicates oscines considered to be Asian (from <http://avibase.bsc-eoc.org>), with relationships described by the tree of ref. 31, including those that Jetz *et al.*³¹ inserted without sequence data. The top row is based on the checklist of Clements 2013 and the second row on Clements 2005, as given in the avibase database. Species with different names in the tree of ref. 31 and the Clements compilations were not included ($N = 178$ for 2005, $N = 467$ for 2013). The lower three rows are examples of trees simulated under the Yule model. In all cases the observed γ statistic for the east Himalayan assemblage lay far outside the range of simulated values: The γ statistic for the phylogeny connecting the 358 oscines in the study area based on the tree used here (Fig. 1b) is −13.94 and from the tree of ref. 31 is −8.30 (357 oscines because two *Corvus* were not split in the tree of ref. 31). The γ statistic for a phylogeny truncated at 3 Myr ago is −14.72 (current tree in Fig. 1b, 346 tips), −10.11 (the tree of ref. 31, 330 tips). The time separating sister pairs in the 1,504 Asian species extracted from the Jetz *et al.* tree³¹ is 4.0 Myr and the average time of separation of sisters in 358 species samples from that tree 4.5 Myr. For further comparisons with the tree of ref. 31 see the phylogeny construction section in the Methods.

Sensory stimulation shifts visual cortex from synchronous to asynchronous states

Andrew Y. Y. Tan^{1,2*}, Yuzhi Chen^{1,2,3*}, Benjamin Scholl^{1,2*}, Eyal Seidemann^{1,2,3} & Nicholas J. Priebe^{1,2}

In the mammalian cerebral cortex, neural responses are highly variable during spontaneous activity and sensory stimulation. To explain this variability, the cortex of alert animals has been proposed to be in an asynchronous high-conductance state in which irregular spiking arises from the convergence of large numbers of uncorrelated excitatory and inhibitory inputs onto individual neurons^{1–4}. Signatures of this state are that a neuron's membrane potential (V_m) hovers just below spike threshold, and its aggregate synaptic input is nearly Gaussian, arising from many uncorrelated inputs^{1–4}. Alternatively, irregular spiking could arise from infrequent correlated input events that elicit large fluctuations in V_m (refs 5, 6). To distinguish between these hypotheses, we developed a technique to perform whole-cell V_m measurements from the cortex of behaving monkeys, focusing on primary visual cortex (V1) of monkeys performing a visual fixation task. Here we show that, contrary to the predictions of an asynchronous state, mean V_m during fixation was far from threshold (14 mV) and spiking was triggered by occasional large spontaneous fluctuations. Distributions of V_m values were skewed beyond that expected for a range of Gaussian input^{6,7}, but were consistent with synaptic input arising from infrequent correlated events^{5,6}. Furthermore, spontaneous fluctuations in V_m were correlated with the surrounding network activity, as reflected in simultaneously recorded nearby local field potential. Visual stimulation, however, led to responses more consistent with an asynchronous state: mean V_m approached threshold, fluctuations became more Gaussian, and correlations between single neurons and the surrounding network were disrupted. These observations show that sensory drive can shift a common cortical circuitry from a synchronous to an asynchronous state.

Cortical neurons show variable activity even after efforts are taken to fix temporal variations in sensory stimuli and attentional state⁸. This ongoing activity affects stimulus encoding and synaptic plasticity⁹, but its neural basis is not well understood. One hypothesis is that the variable activity in alert animals arises from connections between numerous uncorrelated excitatory and inhibitory inputs^{1–4}. Such a network is consistent with studies of neural architecture¹⁰, and shows spiking statistics similar to those measured in extracellular studies⁸. Predictions of this hypothesis^{2–4,6,7} are that numerous uncorrelated inputs (Fig. 1a, bottom) cause V_m to hover near spike threshold (Fig. 1a, top left) and to show distributions that are near Gaussian or skewed with tails at hyperpolarized potentials (Fig. 1a, top right). In contrast, neurons may receive correlated input^{5,6} (Fig. 1b, bottom) such that V_m lies far below threshold and shows infrequent large excursions (Fig. 1b, top left), forming skewed distributions with tails at depolarized potentials (Fig. 1b, top right). Measurements of V_m from awake, non-behaving cats suggest an asynchronous state¹¹, but are also consistent with correlated input¹². Data from behaving rodents in various attentional states have suggested different pictures^{13–16}, but equivocally, because of the potential contributions of uncontrolled sensory inputs and attentional states to V_m dynamics. Extracellular recordings in drowsy humans have demonstrated correlated

spontaneous cortical activity, leaving open the possibility that correlations are absent during alertness¹⁷. Accordingly, we performed the first whole-cell V_m measurements from the cortex of monkeys actively engaged in a visual fixation task, allowing us to examine V_m in single V1 neurons of alert primates while minimizing variability due to sensory stimuli, eye movements and attentional state.

We obtained intracellular¹⁸, whole-cell^{19,20}, current-clamp measurements of V_m from 31 V1 neurons in three macaque monkeys while they viewed gratings of different orientations (see Supplementary Information and Supplementary Video). Each trial began when a fixation spot was displayed at the centre of a monitor in front of the monkey. The monkey had to shift gaze to the fixation point and maintain tight fixation for at least 1,500 ms to receive a reward. A drifting sinusoidal grating was presented for 1,000 ms while the monkey was maintaining strict fixation. We analysed V_m during the fixation period only from trials in which the monkey performed the task successfully. V1 neurons were orientation-selective, and were classified as simple or complex (Supplementary Information and Extended Data Fig. 1).

Comparison of V_m in blank trials in which no visual stimulus was presented (Fig. 2a–c, left) with suprathreshold responses evoked by preferred orientation gratings (Fig. 2a–c, right) shows that blank trial V_m was generally far from spike threshold. There were occasional large depolarizations during blank trials, which were manifested in the positive skewness of V_m amplitude histograms: these had longer tails at depolarized potentials, even though traces had had spikes removed (Fig. 2a–c, left, orange histograms; see also Supplementary Information and Extended Data Fig. 2). Across neurons, the median distance between blank trial

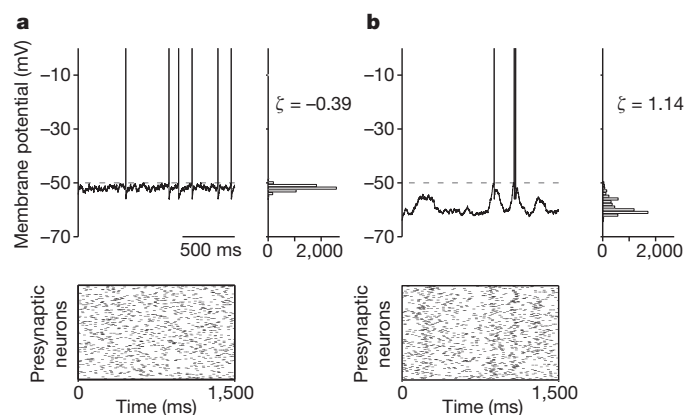


Figure 1 | V_m characteristics depend on network state. **a**, Diagram of an asynchronous high-conductance state. A neuron receives numerous uncorrelated inputs (bottom), V_m hovers near spike threshold (top left), forming distributions with low or negative skewness ζ (top right). **b**, A neuron may instead receive correlated inputs (bottom) such that V_m lies farther from spike threshold and shows occasional large fluctuations (top left), forming distributions with high skewness ζ (top right).

¹Center for Perceptual Systems, University of Texas, Austin, Texas 78712, USA. ²Department of Neuroscience, College of Natural Sciences, University of Texas, Austin, Texas 78712, USA. ³Department of Psychology, University of Texas, Austin, Texas 78712, USA.

*These authors contributed equally to this work.

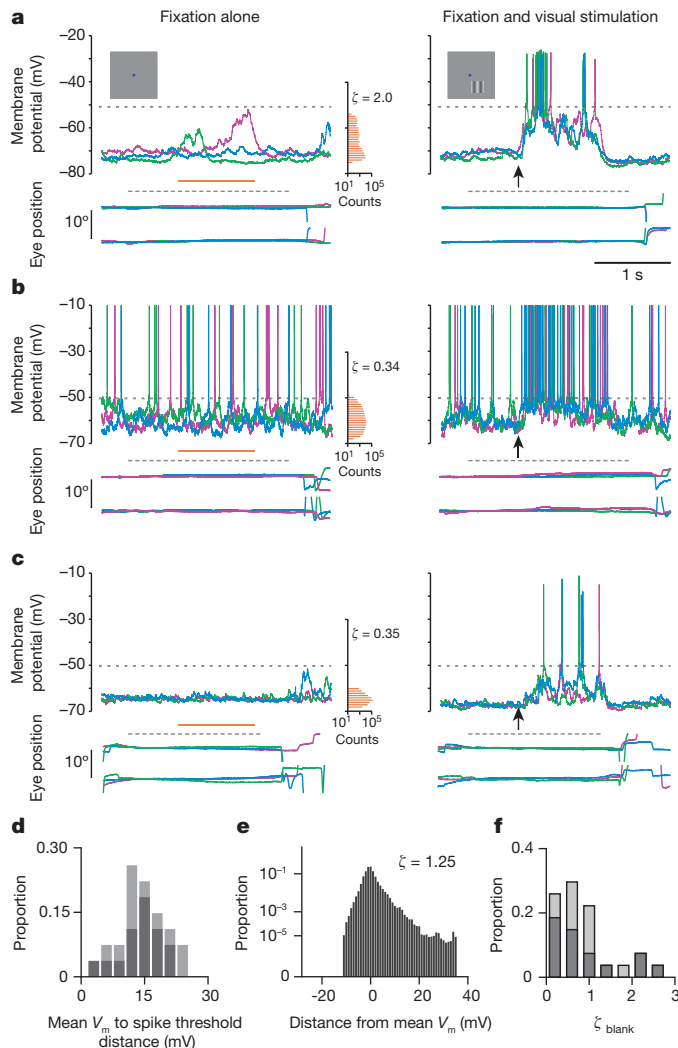


Figure 2 | Occasional large spontaneous fluctuations in V_m during fixation. **a–c**, V_m (top), horizontal and vertical eye position (bottom) from three blank trials, and corresponding histograms from the period indicated by the orange line (left). Horizontal scales for the histograms are logarithmic. Right: traces from three preferred orientation trials; the arrow indicates stimulus onset. Lower and upper dashed lines indicate the period of required fixation and the spike threshold, respectively. Results in **a**, **b** and **c** are from different neurons. **d**, Distribution across neurons of distance between mean V_m during blank trials and spike threshold ($n = 26$). **e**, The population distribution of V_m for blank trials is the average of each neuron's normalized mean-subtracted distribution. **f**, Distribution across neurons of blank trial V_m skewness. Light and dark bars in **d** and **f** indicate simple and complex cells, respectively.

V_m and spike threshold was 13.9 mV (Fig. 2d). The median skewness of 0.72 (Fig. 2e, f) differs from the near zero or negative skewness expected for a range of Gaussian input (Fig. 1a; see also Supplementary Information and Extended Data Fig. 2c), but is consistent with synaptic input arising from infrequent correlated events (Fig. 1b). These data show that in the absence of visual stimulation, V1 of macaques performing a visual fixation task is not in an asynchronous high-conductance state^{1–4}.

By comparison, visual stimulation depolarized neurons (Fig. 2a–c, right, and Fig. 3a–c) and decreased the skewness of V_m deviations from the mean (Fig. 3a–c; see also Supplementary Information and Extended Data Fig. 3), an effect that was significant across the population (Fig. 3d; Wilcoxon signed-rank test, $P < 0.0001$; see also Supplementary Information and Extended Data Fig. 4). Together with observed increases in membrane conductance during visual stimulation^{21,22} (Supplementary Information and Extended Data Fig. 5), these results suggest that

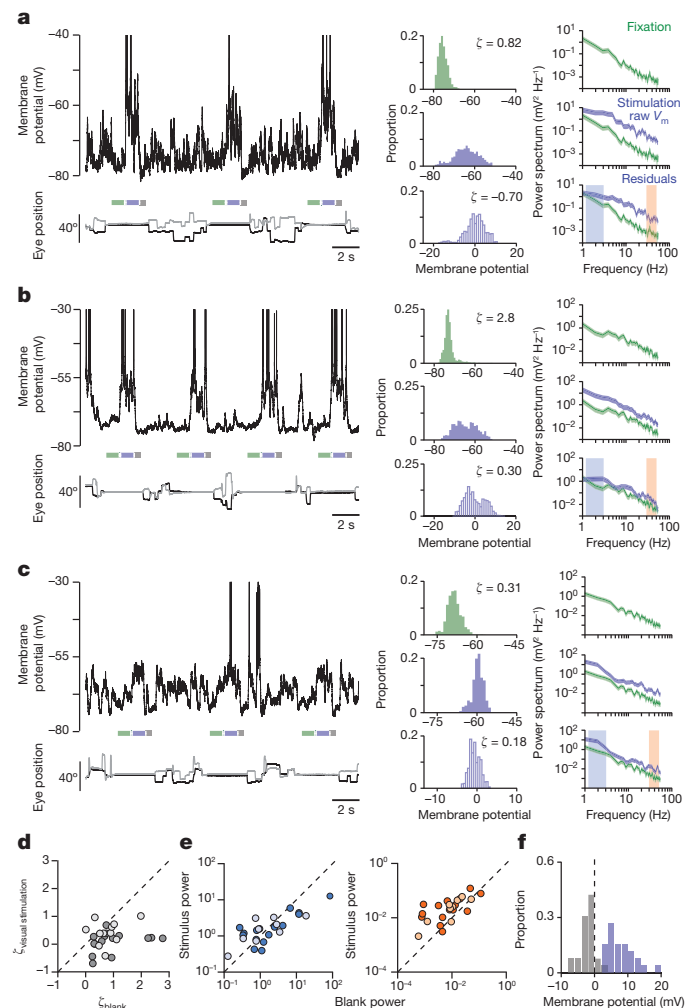


Figure 3 | Visually evoked V_m is closer to threshold and has more Gaussian fluctuations. **a–c**, Left: V_m (top) and eye position (bottom) over pre-stimulus, stimulus and post-stimulus periods during fixation (green, lavender and grey bars, respectively) and inter-trial periods. Right: pre-stimulus, stimulus period raw value and residual (green, lavender-filled and lavender-outlined respectively) histograms and power spectra; shaded areas indicate low and high frequency ranges. Results in **a**, **b** and **c** are from different neurons. **d**, Skewness of V_m residuals during the preferred orientation versus that during blank trials, for each neuron. **e**, Mean power in V_m residuals during blank versus visual stimulation at low frequencies (0.5–4 Hz, left) and high frequencies (30–50 Hz, right). **f**, Distribution across neurons of mean V_m during stimulus (lavender) and post-stimulus (grey) periods, relative to mean pre-stimulus V_m . Light and dark circles in **d** and **e** indicate simple and complex cells, respectively.

visual stimulation shifts the cortical network towards an asynchronous high-conductance state^{1–4}.

Visual stimulation also caused significant changes in the power of V_m fluctuations. Membrane potential showed greater power at low frequencies than at high frequencies during fixation, both before and during visual stimulation. Visual stimulation increased the power of V_m fluctuations from the trial average (that is, residuals) at high frequencies (30–50 Hz) but did not cause systematic changes at low frequencies (0.5–4 Hz) (Fig. 3a–c, e; Wilcoxon signed-rank test, $P = 0.76$ (0.5–4 Hz), $P = 0.001$ (30–50 Hz); see also Supplementary Information and Extended Data Fig. 6). Post-stimulus V_m was typically below pre-stimulus values (Fig. 3f; Wilcoxon signed-rank test $P < 0.0001$).

If, as our intracellular recordings suggest, visual stimulation shifts V1 towards an asynchronous state, there should be a concomitant reduction in the correlation between V_m and the surrounding network, as reflected in the simultaneously recorded nearby local field potential (LFP).

This was the case. During fixation with no visual stimulus, deflections indicating spontaneous increases in activity are evident in V_m and LFP (Fig. 4a, left, depolarization for V_m , downward deflections for LFP). These deflections are coincident in both signals (asterisks in Fig. 4a); across our population, the zero-lag V_m -LFP cross-correlation was negative during blank trials, reflecting coincident activation of the network and individual neurons (Fig. 4d, green, median cross-correlation -0.24 ; Wilcoxon signed-rank test, $P < 0.01$). To determine whether visual stimulation alters this relationship we examined V_m -LFP correlations after trial averages had been subtracted (Fig. 4b, c, centre panels). Correlations declined when drifting gratings were presented (Fig. 4b-d; Wilcoxon signed-rank test, $P < 0.01$), such that the median cross-correlation was nearer zero (Fig. 4d, lavender; Wilcoxon signed-rank test, $P = 0.91$), providing further evidence that visual stimulation drives V1 towards an asynchronous state. The visually-evoked decline in V_m -LFP correlation was apparent for low frequency (0.5–4 Hz), but not high frequency fluctuations (Fig. 4e; Wilcoxon signed-rank test, $P < 0.01$ (0.5–4 Hz), $P = 0.13$ (30–50 Hz)); V_m -LFP coherence decreased at low (0.5–4 Hz), but not high frequencies (30–50 Hz) (Fig. 4b, c, right, and Fig. 4f; Wilcoxon signed-rank test, $P < 0.05$ (0.5–4 Hz), $P = 0.34$ (30–50 Hz); see also Supplementary Information and Extended Data Fig. 7).

We have shown that in the absence of visual stimulation, V1 in alert behaving primates is not in an asynchronous high-conductance state^{1–4}.

Rather, spontaneous V_m fluctuations are non-Gaussian and characterized by occasional excursions from rest, consistent with synaptic input arising from infrequent correlated events^{5,6}. In our recordings, sensory stimulation drove V1 towards an asynchronous state, as visually evoked V_m was closer to spike threshold, showed more Gaussian fluctuations and became less correlated with low-frequency LFP. The visually evoked reduction in correlation between V_m and LFP is consistent with previously reported decreases in spiking correlations^{23,24}. In an analogous fashion, the correlated activity patterns observed in mouse sensory cortex¹⁴ during quiet wakefulness are disrupted by thalamic activation²⁵. (See also Supplementary Information and Extended Data Fig. 8.) Our records focused on activity in superficial cortical layers; membrane potential characteristics may differ across layers, potentially reflecting laminar specificity in network state²⁶.

How can cortical circuitry support synchronous and asynchronous states? One salient difference between the states was the amount of external input: without visual stimulation the thalamic drive to cortex is weak, whereas visual stimulation activates those afferents. We propose that this difference in afferent drive explains the shift in network state. Our proposal unifies observation and theory: a lower input spike rate reduces synaptic input so that V_m lies further from threshold; postsynaptic potentials due to different sources are less likely to overlap in time and instead appear as distinct events. Crucially, theory indicates that a low thalamic

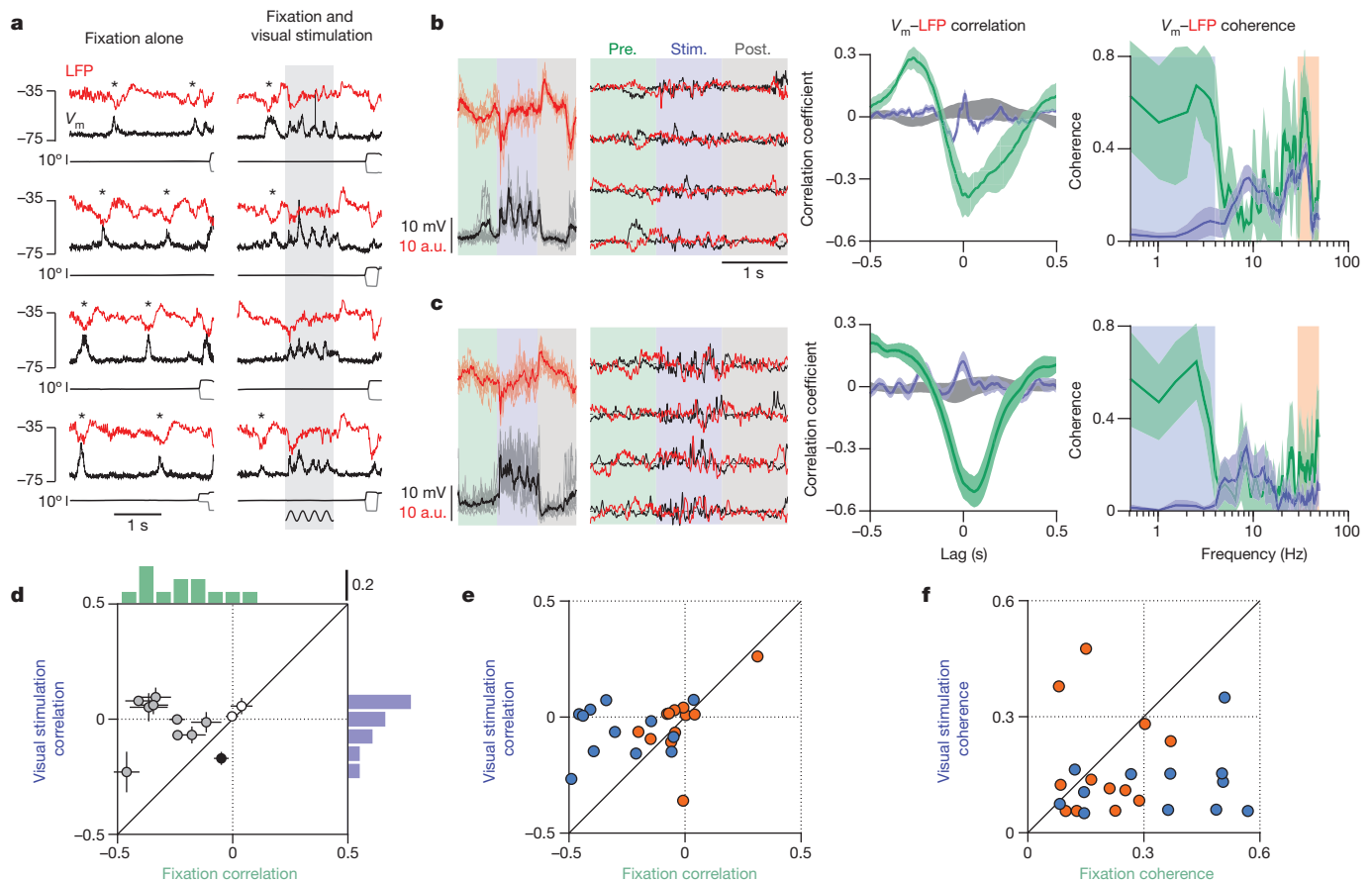


Figure 4 | Magnitude of V_m -LFP cross-correlation decreases during visual stimulation. **a**, Four trials of V_m (black) with simultaneously recorded LFP (red) and eye position (thin black traces) during the blank (left) and the preferred orientation (right) from one neuron. **b**, **c**, Left: V_m and LFP from four trials with superimposed trial average. Centre: traces with trial averages subtracted. Pre., pre-stimulation; stim., stimulation; post., post-stimulation. Right: V_m -LFP cross-correlation functions and coherence magnitudes for blank (green) and stimulus (lavender) periods, and shuffled stimulus trials (grey). Shaded areas show bootstrapped s.e.m. for cross-correlation functions, and jack-knifed 95% confidence intervals for coherence magnitudes. Results in

b and **c** are from different neurons. **d**, Zero-lag V_m -LFP cross-correlation during visual stimulation (lavender) versus blank trials (green) for each neuron, and corresponding marginal histograms. Medium and dark filled circles indicate cross-correlations that were respectively greater or less during visual stimulation than blank trials (Wilcoxon signed-rank test, $P < 0.05$). **e**, Zero-lag V_m -LFP cross-correlation at low frequencies (0.5–4 Hz, blue) and high frequencies (30–50 Hz, orange). **f**, Root-mean-square V_m -LFP coherence magnitudes at low frequencies (0.5–4 Hz, blue) and high frequencies (30–50 Hz, orange).

spike rate destabilizes the asynchronous state towards low-frequency correlations^{4,27,28}, but higher thalamic spike rates drive the network towards an asynchronous state in which correlations weaken^{4,27,28}, as observed in our data. It is clear that external drive alters the cortical state²⁵, but internal factors are also essential. In extrastriate cortex, attention causes an increase in overall response that is also accompanied by a decline in the correlation between neurons^{29,30}. Explaining how these external and internal drives are synthesized will require understanding how V1 interacts with downstream areas.

METHODS SUMMARY

Macaque monkeys were trained to perform a visual fixation task (see also Supplementary Information and Extended Data Fig. 9), and implanted with recording chambers. Blind *in vivo* whole-cell recordings were performed^{19,20}. Patch pipettes were filled with (in mM) 135 potassium gluconate, 4 NaCl, 0.5 EGTA, 2 MgATP²⁻, 10 phosphocreatine disodium, 10 HEPES, pH adjusted to 7.3 with KOH (Sigma-Aldrich) (see also Supplementary Information and Extended Data Fig. 10).

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 27 September 2013; accepted 17 February 2014.

Published online 30 March 2014.

- van Vreeswijk, C. & Sompolinsky, H. Chaos in neuronal networks with balanced excitatory and inhibitory activity. *Science* **274**, 1724–1726 (1996).
- van Vreeswijk, C. & Sompolinsky, H. Chaotic balanced state in a model of cortical circuits. *Neural Comput.* **10**, 1321–1371 (1998).
- Kumar, A., Schrader, S., Aertsen, A. & Rotter, S. The high-conductance state of cortical networks. *Neural Comput.* **20**, 1–43 (2008).
- Renart, A. *et al.* The asynchronous state in cortical circuits. *Science* **327**, 587–590 (2010).
- DeWeese, M. R. & Zador, A. M. Non-Gaussian membrane potential dynamics imply sparse, synchronous activity in auditory cortex. *J. Neurosci.* **26**, 12206–12218 (2006).
- Richardson, M. J. & Gerstner, W. Synaptic shot noise and conductance fluctuations affect the membrane voltage with equal significance. *Neural Comput.* **17**, 923–947 (2005).
- Rudolph, M. & Destexhe, A. Characterization of subthreshold voltage fluctuations in neuronal membranes. *Neural Comput.* **15**, 2577–2618 (2003).
- Tolhurst, D. J., Movshon, J. A. & Thompson, I. D. The dependence of response amplitude and variance of cat visual cortical neurones on stimulus contrast. *Exp. Brain Res.* **41**, 414–419 (1981).
- Legenstein, R., Pecevski, D. & Maass, W. A learning theory for reward-modulated spike-timing-dependent plasticity with application to biofeedback. *PLOS Comput. Biol.* **4**, e1000180 (2008).
- Thomson, A. M. & Lamy, C. Functional maps of neocortical local circuitry. *Front. Neurosci.* **1**, 19–42 (2007).
- Steriade, M., Timofeev, I. & Grenier, F. Natural waking and sleep states: a view from inside neocortical neurons. *J. Neurophysiol.* **85**, 1969–1985 (2001).
- Destexhe, A. & Pare, D. Impact of network activity on the integrative properties of neocortical pyramidal neurons *in vivo*. *J. Neurophysiol.* **81**, 1531–1547 (1999).
- Crochet, S. & Petersen, C. C. Correlating whisker behavior with membrane potential in barrel cortex of awake mice. *Nature Neurosci.* **9**, 608–610 (2006).
- Poulet, J. F. & Petersen, C. C. Internal brain state regulates membrane potential synchrony in barrel cortex of behaving mice. *Nature* **454**, 881–885 (2008).
- Okun, M., Naim, A. & Lampl, I. The subthreshold relation between cortical local field potential and neuronal firing unveiled by intracellular recordings in awake rats. *J. Neurosci.* **30**, 4440–4448 (2010).
- Hromádka, T., Zador, A. M. & Deweese, M. R. Up states are rare in awake auditory cortex. *J. Neurophysiol.* **109**, 1989–1995 (2013).
- Peyrache, A. *et al.* Spatiotemporal dynamics of neocortical excitation and inhibition during human sleep. *Proc. Natl Acad. Sci. USA* **109**, 1731–1736 (2012).
- Matsumura, M., Chen, D., Sawaguchi, T., Kubota, K. & Fetz, E. E. Synaptic interactions between primate precentral cortex neurons revealed by spike-triggered averaging of intracellular membrane potentials *in vivo*. *J. Neurosci.* **16**, 7757–7767 (1996).
- Pei, X., Volgushev, M., Vidyasagar, T. R. & Creutzfeldt, O. D. Whole cell recording and conductance measurements in cat visual cortex *in vivo*. *Neuroreport* **2**, 485–488 (1991).
- Ferster, D. & Jagadeesh, B. EPSP–IPSP interactions in cat visual cortex studied with *in vivo* whole-cell patch recording. *J. Neurosci.* **12**, 1262–1274 (1992).
- Borg-Graham, L. J., Monier, C. & Fregnac, Y. Visual input evokes transient and strong shunting inhibition in visual cortical neurons. *Nature* **393**, 369–373 (1998).
- Hirsch, J. A., Alonso, J. M., Reid, R. C. & Martinez, L. M. Synaptic integration in striate cortical simple cells. *J. Neurosci.* **18**, 9517–9528 (1998).
- Kohn, A. & Smith, M. A. Stimulus dependence of neuronal correlation in primary visual cortex of the macaque. *J. Neurosci.* **25**, 3661–3673 (2005).
- Nauhaus, I., Busse, L., Carandini, M. & Ringach, D. L. Stimulus contrast modulates functional connectivity in visual cortex. *Nature Neurosci.* **12**, 70–76 (2009).
- Poulet, J. F., Fernandez, L. M., Crochet, S. & Petersen, C. C. Thalamic control of cortical states. *Nature Neurosci.* **15**, 370–372 (2012).
- de Kock, C. P. & Sakmann, B. Spiking in primary somatosensory cortex during natural whisking in awake head-restrained rats is cell-type specific. *Proc. Natl Acad. Sci. USA* **106**, 16446–16450 (2009).
- Brunel, N. Dynamics of sparsely connected networks of excitatory and inhibitory spiking neurons. *J. Comput. Neurosci.* **8**, 183–208 (2000).
- Mehring, C., Hehl, U., Kubo, M., Diesmann, M. & Aertsen, A. Activity dynamics and propagation of synchronous spiking in locally connected random networks. *Biol. Cybern.* **88**, 395–408 (2003).
- Cohen, M. R. & Maunsell, J. H. Attention improves performance primarily by reducing interneuronal correlations. *Nature Neurosci.* **12**, 1594–1600 (2009).
- Mitchell, J. F., Sundberg, K. A. & Reynolds, J. H. Spatial attention decorrelates intrinsic activity fluctuations in macaque area V4. *Neuron* **63**, 879–888 (2009).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank T. Cakic for assistance with this project, and J. Hanover, D. Ferster, K. D. Miller and A. C. Huk for discussions and comments. A.Y.Y.T., B.S. and N.J.P. were supported by grants from the National Institutes of Health (NIH) (EY-019288) and the Pew Charitable Trusts; Y.C. and E.S. were supported by grants from the NIH (EY-016454 and EY-16752).

Author Contributions A.Y.Y.T., E.S. and N.J.P. initiated and designed the study. All authors collected the data, analysed the results, discussed the findings and wrote the paper. A.Y.Y.T., Y.C. and B.S. contributed equally to this work. E.S. and N.P. contributed equally to this work.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to A.Y.Y.T. (aty@alum.mit.edu) or N.J.P. (nico@austin.utexas.edu).

METHODS

All procedures were approved by the University of Texas Institutional Animal Care and Use Committee and conformed to National Institutes of Health standards. Our general experimental procedures in behaving macaque monkeys have previously been described in detail^{31,32}.

Behavioural task and visual stimulus. Three adult male macaque monkeys (*Macaca mulatta*) were trained to perform a visual fixation task in which gratings of different orientations were presented. Each trial began when a fixation spot was displayed at the centre of a monitor in front of the monkey. The monkey had to shift gaze to the fixation point and maintain fixation within a small window (less than 2° full width) for at least 1,500 ms to receive a reward. A drifting sinusoidal grating was presented at a randomized orientation for 1,000 ms while the monkey was maintaining strict fixation, thus minimizing variability due to eye movements. (See Supplementary Information and Extended Data Fig. 9 for characteristics of post-fixation saccades.)

Visual stimuli were presented on a gamma-corrected high-end 21-inch colour display (Sony Trinitron GDM-F520) at a fixed mean luminance of 30 cd m⁻². The display subtended 20.5° × 15.4° at a viewing distance of 108 cm and had a pixel resolution of 1,024 × 768, 30-bit colour depth and a refresh rate of 100 Hz. Visual stimuli were generated by using a high-end graphics card on a dedicated PC, using custom-designed software. Behavioural measurements and data acquisition were controlled by a PC running a software package for neurophysiological recordings from alert animals (Reflective Computing). Eye movements were measured with an infrared eye-tracking device (Dr Bouis).

Whole cell recordings. Recording chambers were located on the dorsal portion of V1, with the anterior portion of the chamber reaching close to the lunate sulcus and the border between V1 and V2. We verified the retinotopic organization by voltage-sensitive dye imaging³³, and by recording multiunit activity or local field potential with tungsten microelectrodes (Alpha Omega Co; MicroProbes for Life Sciences). The cortex in our cranial windows represents stimuli that are approximately 2.5°–5° away from the fovea in the lower quadrant of the contralateral hemifield.

Intracellular records of V_m (refs 18, 34, 35) were obtained with blind *in vivo* whole-cell recordings^{19–22}. The recording chamber was filled with 2–4% agarose in artificial cerebrospinal fluid (CSF). Intracellular records were from neurons in the top 1,300 μm of V1. As a reference electrode, a silver–silver chloride wire was inserted into the agarose. The potential of the CSF was assumed to be uniform and equal to that of the reference electrode. Pipettes (6–12 MΩ) were pulled from borosilicate glass capillaries (KG-33, 1.2 mm outer diameter, 0.70 mm inner diameter; King Precision Glass) on a P-2000 micropipette puller (Sutter Instruments). Patch pipettes were filled with (in mM) 135 potassium gluconate, 4 NaCl, 0.5 EGTA, 2 MgATP²⁻, 10 phosphocreatine disodium, 10 HEPES, pH adjusted to 7.3 with KOH (Sigma–Aldrich). Whole-cell current-clamp recordings were performed with an Axoclamp 2B Microelectrode Amplifier (Molecular Devices). We subtracted 7 mV from all raw membrane potential values to compensate for the liquid junction potential³⁶. (See Supplementary Information and Extended Data Fig. 10 for intrinsic properties of recorded neurons.)

Data analysis. We analysed V_m during the fixation period in trials during which the monkey performed the task successfully, provided that the mean V_m in the absence of a visual stimulus was less than –50 mV. V_m was detrended by high-pass filtering at 0.1 Hz. Data were analysed with MATLAB (Mathworks). Shot noise contributions to V_m were assessed by the skewness^{5,37–39} of V_m distributions. Coherence estimates were performed with Chronux⁴⁰, a MATLAB library (freely available from <http://chronux.org/>).

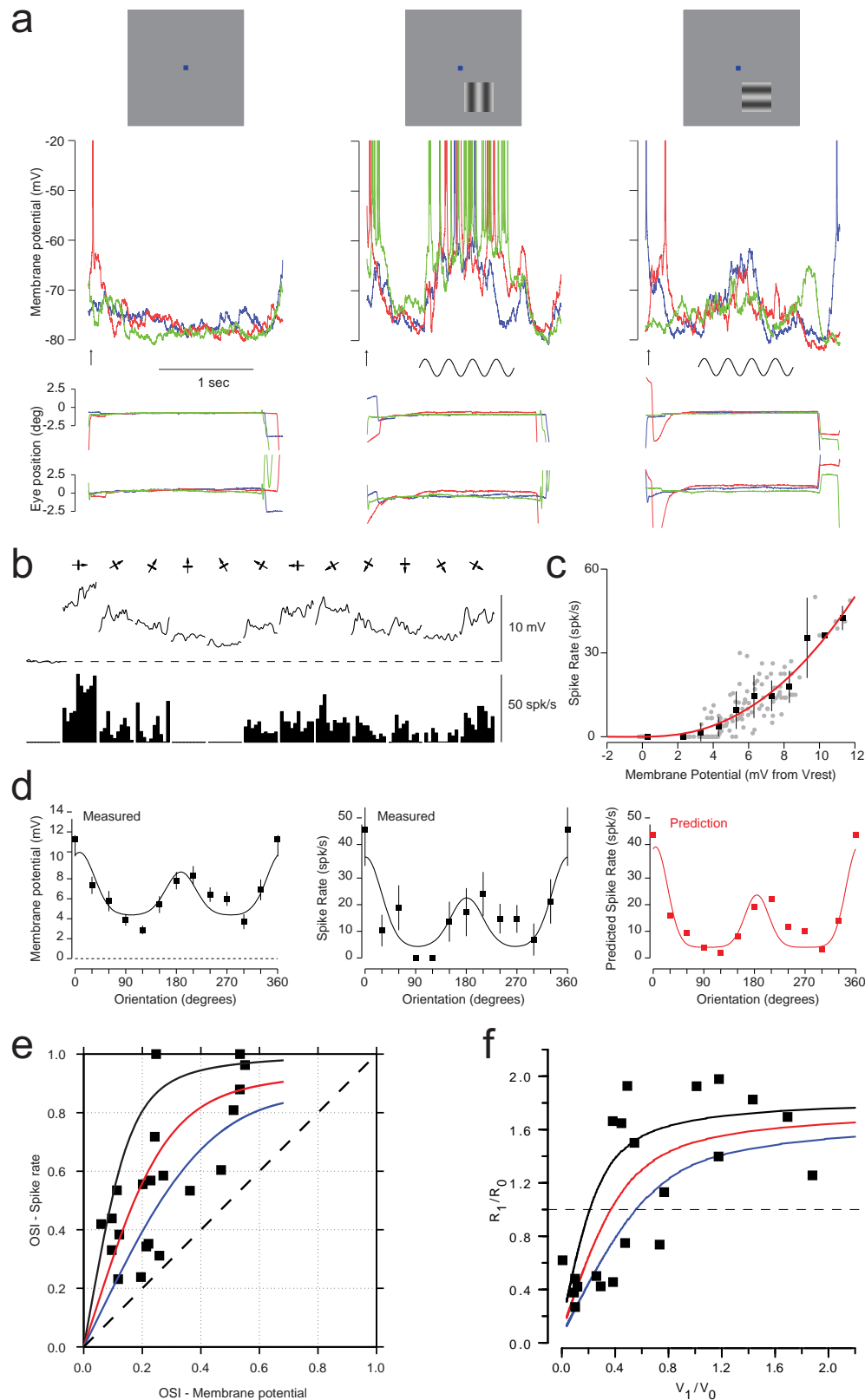
Data analysis for Supplementary Information. The relationship between spike rate R and V_m was described with a threshold followed by a power law^{41–43}: $R = A[V_m - V_{th}]_+^p$, where A is a fitted constant, V_{th} is the resting membrane potential, $+$ indicates rectification, and p is the fitted exponent. Orientation selectivity was

assessed with an orientation selectivity index^{44,45} (vector average = 1 – circular variance). Temporal modulation was assessed with the Fourier component of the response with the same temporal frequency as the moving sinusoidal grating visual stimulus divided by the time-averaged response⁴⁶ (F_1/F_0). Simulations of Hodgkin–Huxley neurons used parameters adapted from refs 47 and 48, and were performed with Brian^{49,50}. We estimated membrane conductance from voltage responses to hyperpolarizing current pulses of constant amplitude, and a fit of a sum of two exponentials to the voltage response⁵¹:

$$V(t) = I_{inj}[(R_M(1 - \exp(-t/\tau_M)) + (R_E(1 - \exp(-t/\tau_E)))]$$

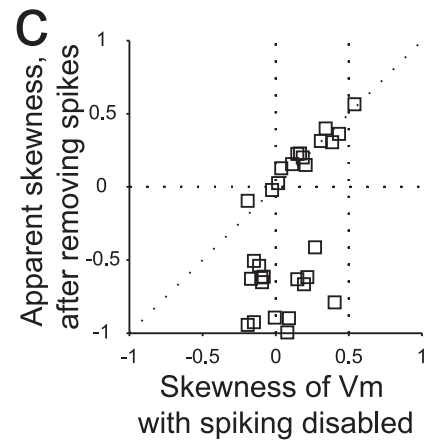
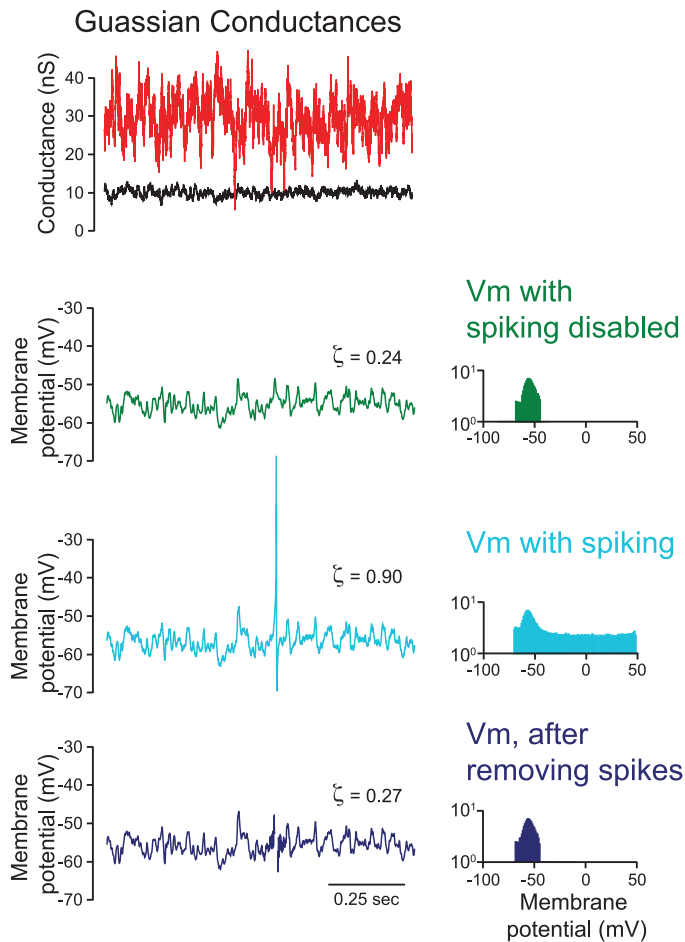
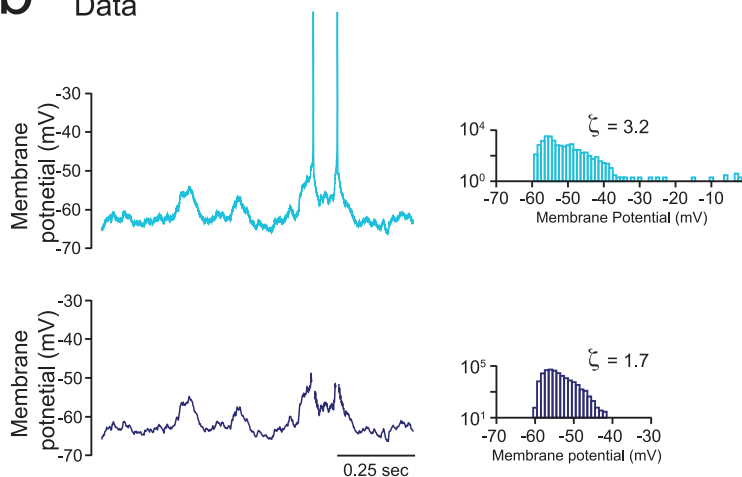
where V is the voltage response, t is time, I_{inj} is injected current, R_M is membrane resistance, τ_M is membrane time constant, R_E is electrode resistance and τ_E is electrode time constant. Membrane conductance is $1/R_M$.

31. Chen, Y., Geisler, W. S. & Seidemann, E. Optimal decoding of correlated neural population responses in the primate visual cortex. *Nature Neurosci.* **9**, 1412–1420 (2006).
32. Chen, Y., Geisler, W. S. & Seidemann, E. Optimal temporal decoding of neural population responses in a reaction-time visual detection task. *J. Neurophysiol.* **99**, 1366–1379 (2008).
33. Yang, Z., Heeger, D. J. & Seidemann, E. Rapid and precise retinotopic mapping of the visual cortex obtained by voltage-sensitive dye imaging in the behaving monkey. *J. Neurophysiol.* **98**, 1002–1014 (2007).
34. Matsumura, M. Intracellular synaptic potentials of primate motor cortex neurons during voluntary movement. *Brain Res.* **163**, 33–48 (1979).
35. Chen, D. & Fetis, E. E. Characteristic membrane potential trajectories in primate sensorimotor cortex neurons recorded *in vivo*. *J. Neurophysiol.* **94**, 2713–2725 (2005).
36. Margrie, T. W., Brecht, M. & Sakmann, B. *In vivo*, low-resistance, whole-cell recordings from neurons in the anaesthetized and awake mammalian brain. *Pflügers Arch.* **444**, 491–498 (2002).
37. Richardson, M. J. & Gerstner, W. Statistics of subthreshold neuronal voltage fluctuations due to conductance-based synaptic shot noise. *Chaos* **16**, 026106 (2006).
38. Richardson, M. J. & Swadlow, R. Firing-rate response of a neuron receiving excitatory and inhibitory synaptic shot noise. *Phys. Rev. Lett.* **105**, 178102 (2010).
39. Wolff, L. & Lindner, B. Method to calculate the moments of the membrane voltage in a model neuron driven by multiplicative filtered shot noise. *Phys. Rev. E* **77**, 041913 (2008).
40. Mitra, P. & Bokil, H. *Observed Brain Dynamics* (Oxford Univ. Press, 2008).
41. Miller, K. D. & Troyer, T. W. Neural noise can explain expansive, power-law nonlinearities in neural response functions. *J. Neurophysiol.* **87**, 653–659 (2002).
42. Hansel, D. & van Vreeswijk, C. How noise contributes to contrast invariance of orientation tuning in cat visual cortex. *J. Neurosci.* **22**, 5118–5128 (2002).
43. Priebe, N. J., Mechler, F., Carandini, M. & Ferster, D. The contribution of spike threshold to the dichotomy of cortical simple and complex cells. *Nature Neurosci.* **7**, 1113–1122 (2004).
44. Swindale, N. V. Orientation tuning curves: empirical description and estimation of parameters. *Biol. Cybern.* **78**, 45–56 (1998).
45. Ringach, D. L., Shapley, R. M. & Hawken, M. J. Orientation selectivity in macaque V1: diversity and laminar dependence. *J. Neurosci.* **22**, 5639–5651 (2002).
46. Skottun, B. C. *et al.* Classifying simple and complex cells on the basis of response modulation. *Vision Res.* **31**, 1079–1086 (1991).
47. Destexhe, A., Rudolph, M., Fellous, J. M. & Sejnowski, T. J. Fluctuating synaptic conductances recreate *in vivo*-like activity in neocortical neurons. *Neuroscience* **107**, 13–24 (2001).
48. Pospisil, M. *et al.* Minimal Hodgkin–Huxley type models for different classes of cortical and thalamic neurons. *Biol. Cybern.* **99**, 427–441 (2008).
49. Goodman, D. & Brette, R. Brian: a simulator for spiking neural networks in python. *Front. Neuroinform.* **2**, 5 (2008).
50. Goodman, D. F. & Brette, R. The Brian simulator. *Front. Neurosci.* **3**, 192–197 (2009).
51. Anderson, J. S., Carandini, M. & Ferster, D. Orientation tuning of input conductance, excitation, and inhibition in cat primary visual cortex. *J. Neurophysiol.* **84**, 909–926 (2000).



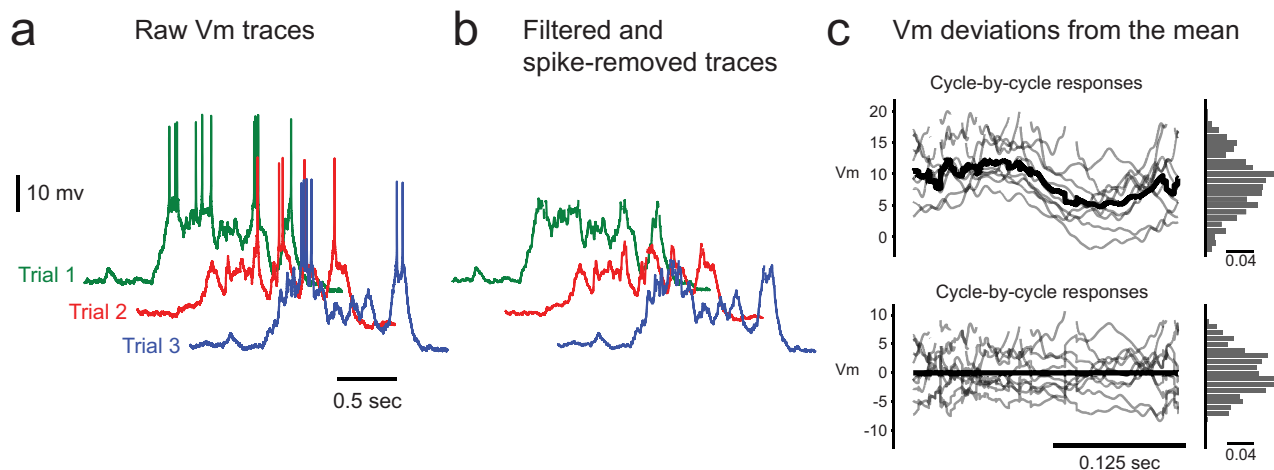
Extended Data Figure 1 | Orientation tuning of V_m and spike rate. **a**, V_m responses (top traces), eye position traces (bottom pairs of traces) from three blank trials (left), three trials at the preferred orientation (centre), and three trials at the orthogonal orientation (right). **b**, Trial averaged V_m (top) and spike rate (bottom) for all orientations, from the neuron in **a**. **c**, Spike rate versus membrane potential, and best-fit thresholded power law, from the neuron in **a**. **d**, Orientation tuning curves for V_m and spike rate, and predicted spike rate orientation tuning curve using the V_m orientation tuning curve and the best-fit thresholded power law in **c**, from the neuron in **a**. **e**, Orientation selectivity

index (OSI) for spike rate versus OSI for V_m . Lines represent expected relationships between spike rate OSI and V_m OSI for thresholded power laws with exponents 2, 3 and 5 (blue, red and black, respectively). **f**, Fourier component of the response with the same temporal frequency as the moving sinusoidal grating visual stimulus divided by the time-averaged response for spike rate (R_1/R_0) versus that for V_m (V_1/V_0). Lines represent expected relationships between R_1/R_0 and V_1/V_0 for thresholded power laws with exponents 2, 3 and 5 (blue, red and black, respectively).

a Simulated Data**b** Data

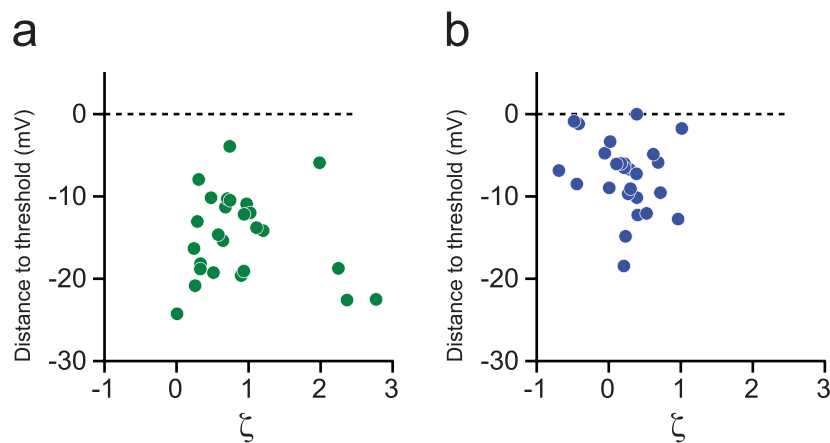
Extended Data Figure 2 | Estimation and implication of V_m skewness during blank trials. **a**, Gaussian excitatory (black) and inhibitory (red) conductances, V_m with spiking disabled (green), V_m with spiking enabled (light blue), and V_m with spikes removed (dark blue), and corresponding V_m amplitude histograms and skewness values ζ , for a simulated neuron with Hodgkin–Huxley conductances. **b**, V_m with spiking (light blue) and with spikes

removed (dark blue) and corresponding V_m amplitude histograms and skewness values ζ , for a recorded neuron. **c**, Apparent skewness from V_m with spikes removed versus skewness from V_m with spiking disabled from a simulated neuron with Hodgkin–Huxley conductances, for a range of Gaussian inputs.



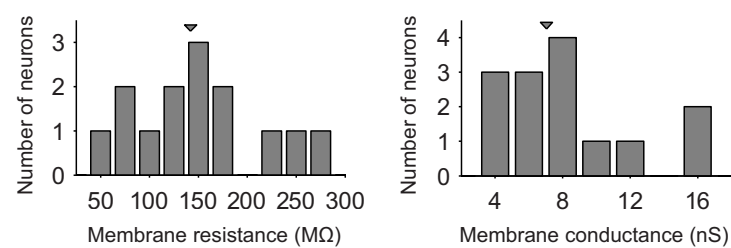
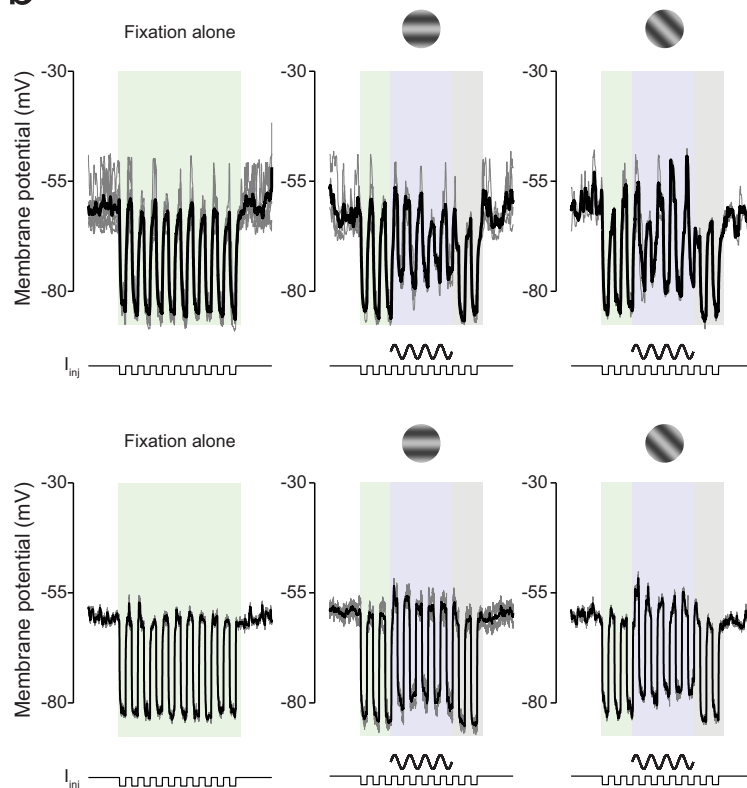
Extended Data Figure 3 | Estimation of V_m skewness during visual stimulation trials. **a**, Raw traces from several trials. **b**, Traces after bandpass filtering and spike removal. **c**, V_m responses from each cycle (top grey traces), cycle-averaged response (top black trace) and histogram of V_m responses (top

histogram); residual traces from each cycle after subtraction of cycle-averaged response (bottom grey traces), cycle-averaged residuals (bottom black trace) and histogram of V_m residuals (bottom histogram). Note the change in vertical scale from top to bottom panels.



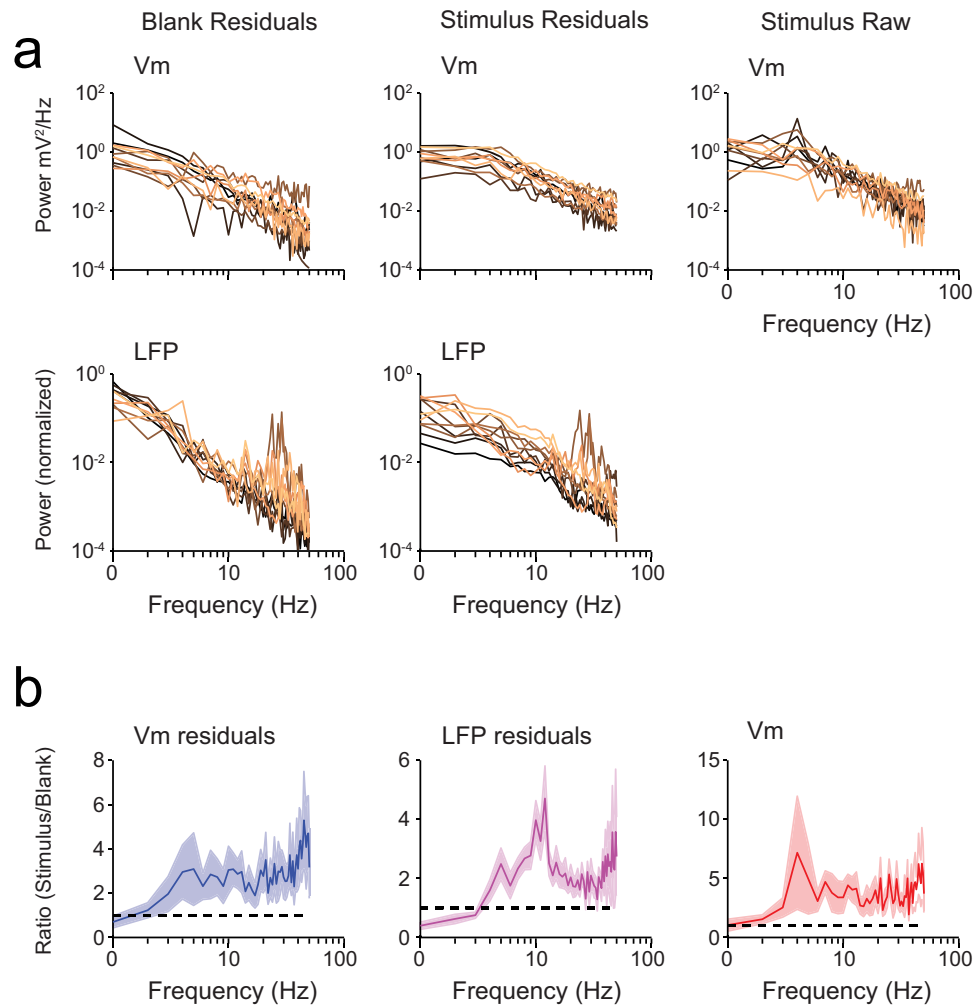
Extended Data Figure 4 | Joint distribution of V_m -threshold distance and skewness. **a**, Joint distribution of V_m -threshold distance and skewness ζ during

blank trials. **b**, Joint distribution of V_m -threshold distance and skewness ζ during preferred orientation trials.

a**b**

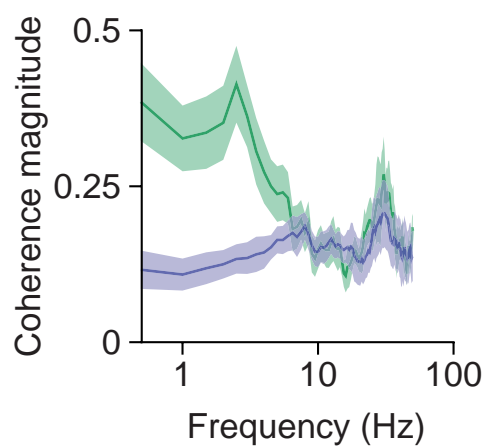
Extended Data Figure 5 | Membrane conductance during blank and visual stimulation trials. **a**, Distribution of membrane resistance (left) and corresponding membrane conductance (right) during blank trials. **b**, Change in

membrane conductance during visual stimulation in two example neurons during blank (left), preferred (centre) and 45° from preferred (right) trials. Each row shows data from a different neuron.

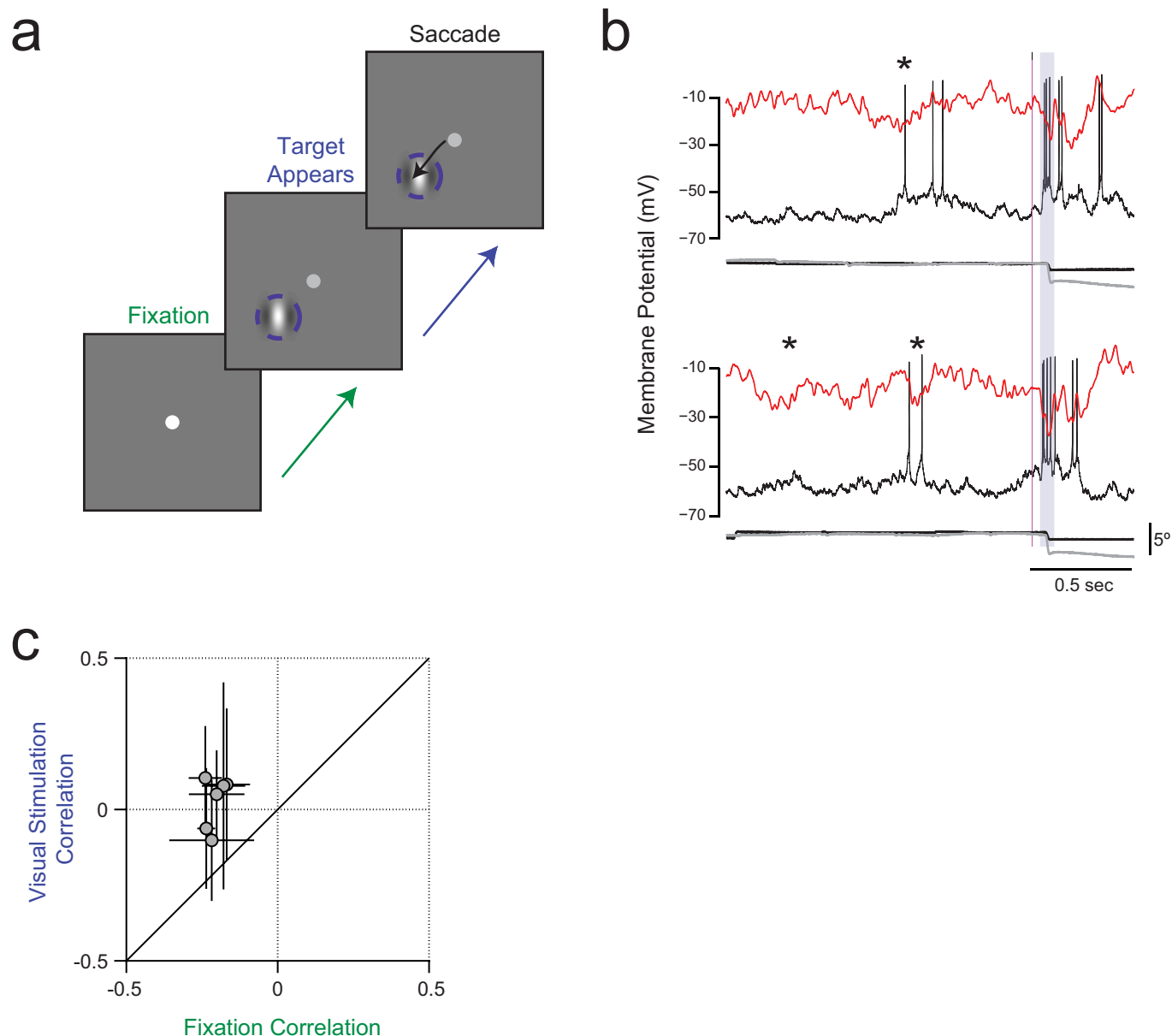


Extended Data Figure 6 | Power spectra of V_m and LFP fluctuations from the trial average. **a**, Power spectrum of V_m (top panels) and LFP (bottom panels) fluctuations from the trial average (residuals) during blank trials (left panels), residuals during preferred orientation stimulation (middle panels), and raw V_m traces during preferred orientation stimulation (right

panel). Each trace corresponds to an individual neuron. **b**, Population-averaged ratio of power spectrum at the preferred orientation to power spectrum for blank trials for V_m fluctuations from the trial average ('V_m residuals', left panel), LFP fluctuations from the trial average ('LFP residuals', middle panel), and raw V_m (right panel). Error bars are jack-knifed standard errors.

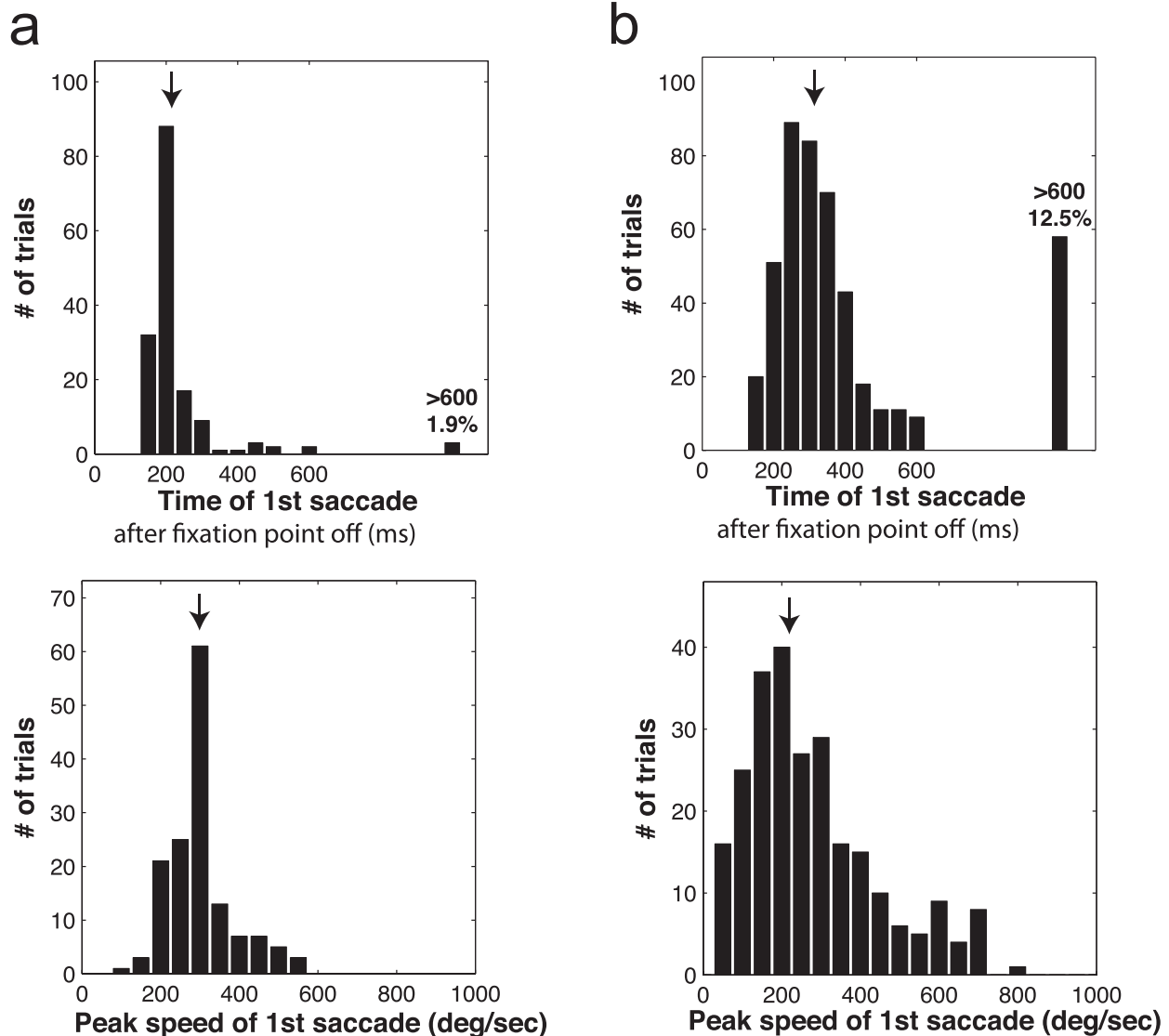


Extended Data Figure 7 | V_m -LFP coherence magnitude for blank trials and visual stimulation. Population-averaged V_m -LFP coherence magnitudes for blank trials (green) and at the preferred orientation (lavender). Error bars are jack-knifed standard errors.



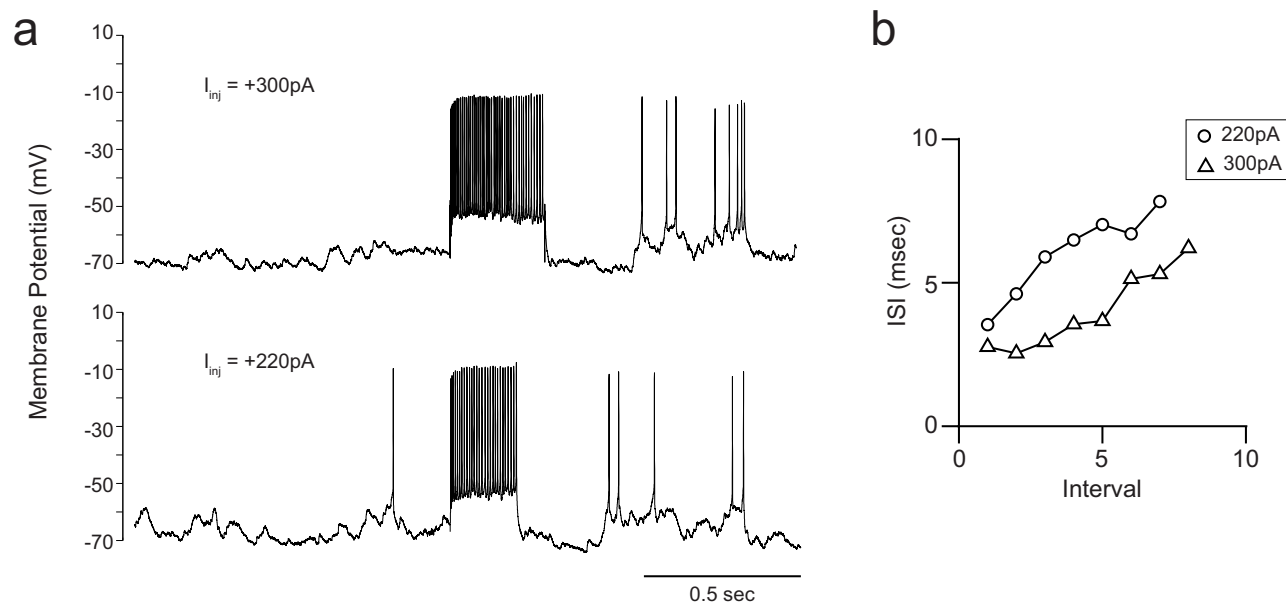
Extended Data Figure 8 | Decreased magnitude of V_m -LFP correlation during a flashed stimulus in a visual saccade task. **a**, Each trial began when a fixation spot was displayed at the centre of a monitor in front of the monkey. The monkey had to shift gaze to the fixation point and maintain tight fixation for at least 1,500 ms. A flashed Gabor target stimulus appeared at a random time between 1,000 and 1,500 ms after the monkey had established tight fixation. The monkey had to saccade to the target stimulus within 600 ms to receive a reward. We analysed V_m and LFP only from trials in which the monkey performed the task successfully. **b**, Simultaneously recorded V_m and

LFP, as well as eye movement traces, in two trials from an example neuron. Asterisks indicate near-simultaneous deflections in V_m and LFP during the pre-stimulus fixation period. Grey shading indicates the analysis period for correlations during the flashed Gabor stimulus; we included 30 ms after saccade onset in this period, because the visual latency for spike responses in the lateral geniculate nucleus is greater than 30 ms. **c**, Zero-lag cross-correlation between V_m and LFP fluctuations from the trial average during the flashed Gabor stimulus versus during the pre-stimulus period.



Extended Data Figure 9 | Summary of first saccade latency and peak velocity in monkeys T and W, which together contributed the majority of the recorded data. a, Top: histogram of latency of first saccade after fixation point termination in three neurons (158 trials) in monkey W. Arrow indicates median latency (217 ms). In 1.9% of the trials no saccade was detected in the 600 ms after fixation point termination. **Bottom:** histogram of peak eye velocity for first saccades during the 600 ms after fixation point offset. Arrow indicates median peak velocity (292° s⁻¹). **b, Results from eight neurons (464 trials) in monkey T.** The format is the same as in **a**. Median latency is

314 ms and median peak velocity is 229° s⁻¹. Monkey W tended to make larger saccades away from fixation, whereas monkey T tended to make smaller saccades and in a small subset of the trials remained close to the fixation point location until the next trial was initiated. This may reflect the fact that the minimal inter-trial interval was shorter in monkey T than in monkey W. The short latency of the saccades after fixation point termination in the vast majority of the trials indicates that both monkeys were alert and attentive and were actively engaged in maintaining tight fixation.



Extended Data Figure 10 | Regular-spiking neurons. **a**, V_m response to injections of current steps of different magnitudes in an example neuron. **b**, Interspike interval during the current step versus interval ordinal. The

interspike interval increased with interval ordinal, indicating that this neuron was regular-spiking.

Listeria monocytogenes exploits efferocytosis to promote cell-to-cell spread

Mark A. Czuczman^{1,2}, Ramzi Fattouh¹, Jorik M. van Rijn³, Veronica Canadien¹, Suzanne Osborne¹, Aleixo M. Muijs^{1,4,5,6}, Vijay K. Kuchroo⁷, Darren E. Higgins⁸ & John H. Brumell^{1,2,5,6}

Efferocytosis, the process by which dying or dead cells are removed by phagocytosis, has an important role in development, tissue homeostasis and innate immunity¹. Efferocytosis is mediated, in part, by receptors that bind to exofacial phosphatidylserine (PS) on cells or cellular debris after loss of plasma membrane asymmetry. Here we show that a bacterial pathogen, *Listeria monocytogenes*, can exploit efferocytosis to promote cell-to-cell spread during infection. These bacteria can escape the phagosome in host cells by using the pore-forming toxin listeriolysin O (LLO) and two phospholipase C enzymes². Expression of the cell surface protein ActA allows *L. monocytogenes* to activate host actin regulatory factors and undergo actin-based motility in the cytosol, eventually leading to formation of actin-rich protrusions at the cell surface. Here we show that protrusion formation is associated with plasma membrane damage due to LLO's pore-forming activity. LLO also promotes the release of bacteria-containing protrusions from the host cell, generating membrane-derived vesicles with exofacial PS. The PS-binding receptor TIM-4 (encoded by the *Timd4* gene) contributes to efficient cell-to-cell spread by *L. monocytogenes* in macrophages *in vitro* and growth of these bacteria is impaired in *Timd4*^{-/-} mice. Thus, *L. monocytogenes* promotes its dissemination in a host by exploiting efferocytosis. Our results indicate that PS-targeted therapeutics may be useful in the fight against infections by *L. monocytogenes* and other bacteria that use similar strategies of cell-to-cell spread during infection.

The intermediate stages of cell-to-cell spread by *L. monocytogenes* remain unclear. On the basis of observations with an *in vitro* infection model, it has been suggested that bacteria-containing protrusions are released from infected cells before uptake of membrane vesicles containing bacteria by neighbouring cells³. However, the mechanisms that mediate protrusion release and uptake of bacteria in vesicles are not known.

LLO is required for *L. monocytogenes* cell-to-cell spread in some cell types, including macrophages^{4,5}. LLO is a pore-forming toxin that is often referred to as a 'phagosome-specific lysin'⁶ because it has limited activity in the cytosol of host cells, owing to its relatively low lytic activity⁷ and stability⁸ at neutral pH. Furthermore, LLO is degraded by the proteasome⁹. Despite these factors, it is now appreciated that LLO can damage the plasma membrane of host cells¹⁰. Host membrane repair pathways limit LLO-mediated membrane damage¹¹, but the mechanisms by which they act remain unclear. LLO is essential for disruption of the outer membrane of spreading vacuoles⁴. Whether LLO contributes to other stages of cell-to-cell spread has not been tested.

We proposed that LLO-mediated damage to the plasma membrane may promote cell-to-cell spread. We used a propidium iodide (PI) assay to measure membrane damage induced during *L. monocytogenes* infection (Fig. 1a). Repair of the plasma membrane is a Ca²⁺-dependent process¹². Therefore, the absence of Ca²⁺ in the medium provided a convenient method to inactivate endogenous repair mechanisms and visualize the full extent of membrane damage. HeLa cells were used for

these studies as phagosome escape by *L. monocytogenes* does not require LLO in this cell type¹³.

In the absence of extracellular Ca²⁺, infection of cells with wild-type bacteria revealed an increase in membrane damage compared with uninfected cells (Fig. 1b, c). The number of PI⁺ cells increased over time, indicating that membrane damage was an ongoing event during infection. Less damage was observed when Ca²⁺ was present in the extracellular medium, suggesting that Ca²⁺-dependent repair pathways limit plasma membrane damage.

Caspase 7 promotes membrane repair during *L. monocytogenes* infection of macrophages¹¹. Consistent with this, we found that short interfering RNA (siRNA)-mediated knockdown of caspase 7 increased membrane damage induced by *L. monocytogenes* (Extended Data Fig. 1a, b). However, this effect was minor, indicating that other factors contribute to membrane repair. Annexins also have a role in membrane repair¹⁴. We found that siRNA-mediated knockdown of annexins A1, A2 and A6 led to an increase in membrane damage (Extended Data Fig. 1a, b). We conclude that multiple host factors contribute to repair of the plasma membrane during *L. monocytogenes* infection.

LLO damages host membranes during infection^{10,11}. Consistent with this, a mutant lacking LLO (*Δhly*) was impaired in membrane damage, and complementation with *hly* restored membrane damage (Fig. 1d and Extended Data Fig. 2a). Deletion of both phospholipase C (PLC) enzymes had no effect on membrane damage in Ca²⁺-free media. However, PLCs were required for membrane damage in Ca²⁺-containing media, suggesting that they may promote LLO activity and/or inhibit membrane repair mechanisms. We observed a decrease in membrane damage in cells infected with ActA-deficient (*ΔactA*) bacteria. Treatment of cells with the actin inhibitors latrunculin B or cytochalasin D had a similar effect (Extended Data Fig. 2b). Thus, actin-based motility may promote membrane damage by LLO by allowing close apposition of bacteria to the plasma membrane.

Membrane damage is associated with the loss of membrane asymmetry and redistribution of PS to the outer leaflet of the plasma membrane¹⁵. Therefore, to visualize membrane damage induced by *L. monocytogenes* we stained cells with a probe (annexin A5–Alexa 488) to label exofacial PS. In uninfected cells, low amounts of exofacial PS were detected (Extended Data Fig. 3). By contrast, treatment of cells with saponin led to staining of cells with annexin A5–Alexa 488.

In cells infected by wild-type bacteria, we observed the formation of PS⁺ structures at the cell surface (Fig. 2a, right, and Extended Data Fig. 4). These structures co-localized with bacteria (visualized by red fluorescent protein (RFP) expression¹⁶) and were not associated with cellular blebbing, rounding or nuclear condensation, indicating that they were not the consequence of apoptosis. Extracellular bacteria did not label with annexin A5–Alexa 488 and bacteria present in PS⁺ structures were not accessible to anti-*L. monocytogenes* antibodies (Extended Data Fig. 5). This indicated that PS⁺ bacteria were present in a host-derived membrane

¹Cell Biology Program, Hospital for Sick Children, Toronto, Ontario M5G0A4, Canada. ²Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S1A8, Canada. ³Department of Cell Biology and Institute of Biomembranes, University Medical Center Utrecht, 3584 CX Utrecht, the Netherlands. ⁴Division of Gastroenterology, Hepatology, and Nutrition, Department of Paediatrics, Hospital for Sick Children, Toronto, Ontario M5G1X8, Canada. ⁵Institute of Medical Science, University of Toronto, Toronto, Ontario M5S1A8, Canada. ⁶Sickkids IBD Centre, Hospital for Sick Children, Toronto, Ontario M5G1X8, Canada. ⁷Center for Neurologic Diseases, Brigham and Women's Hospital, Harvard Medical School, Boston, Massachusetts 02115, USA. ⁸Department of Microbiology and Immunobiology, Harvard Medical School, Boston, Massachusetts 02115, USA.

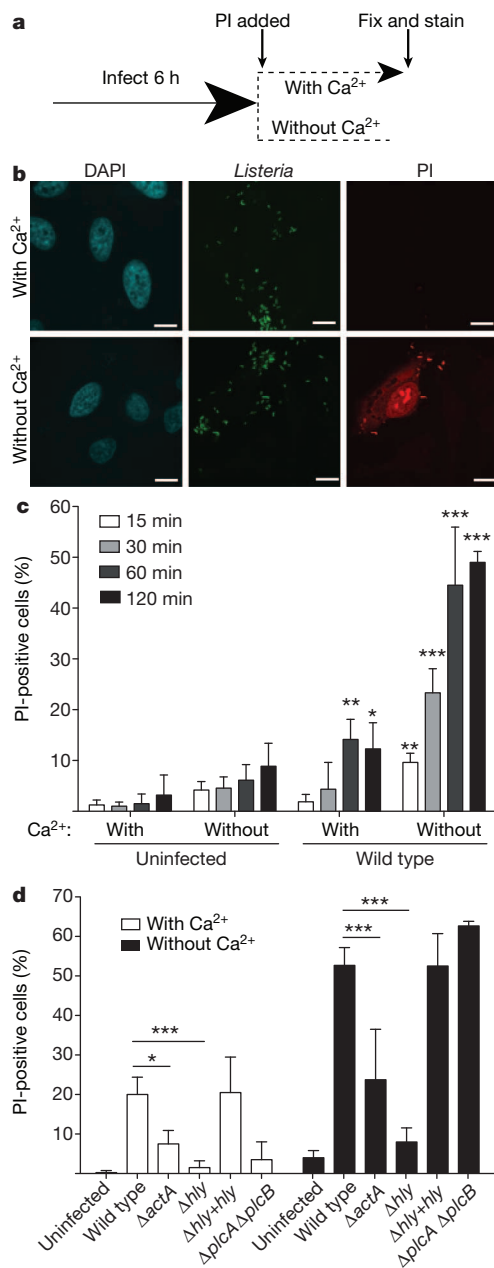


Figure 1 | Actin-based motility promotes LLO-mediated membrane damage. **a**, Experimental design for membrane damage assay. **b**, Confocal images of HeLa cells infected as in **a** with wild-type *L. monocytogenes* at a multiplicity of infection (MOI) of 100. DAPI, 4',6-diamidino-2-phenylindole. Scale bars, 10 μm . Images are representative of three independent experiments. **c**, Cells were infected as in **a** for the indicated time and PI⁺ cells were enumerated ($n = 100$). Data are shown as averages \pm standard deviation (s.d.) for four independent experiments. P values were calculated using two-tailed Student's t -test. **d**, Averages \pm s.d. for three independent experiments are shown. P values were calculated using one-way analysis of variance (ANOVA). * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$.

structure. Correlative microscopy revealed that the PS⁺ structures were associated with the surface of infected cells (Fig. 2a, left). Formation of PS⁺ structures required LLO, consistent with a role for membrane damage in their formation (Fig. 2b, c). PS⁺ structures were also observed in Henle-407 cells and murine bone-marrow-derived macrophages (BMDMs) (Extended Data Fig. 6).

Several lines of evidence indicated that PS⁺ structures are associated with protrusion formation. First, we observed filamentous PS⁺ structures connected to bacteria that resembled protrusions (Fig. 2b, top).

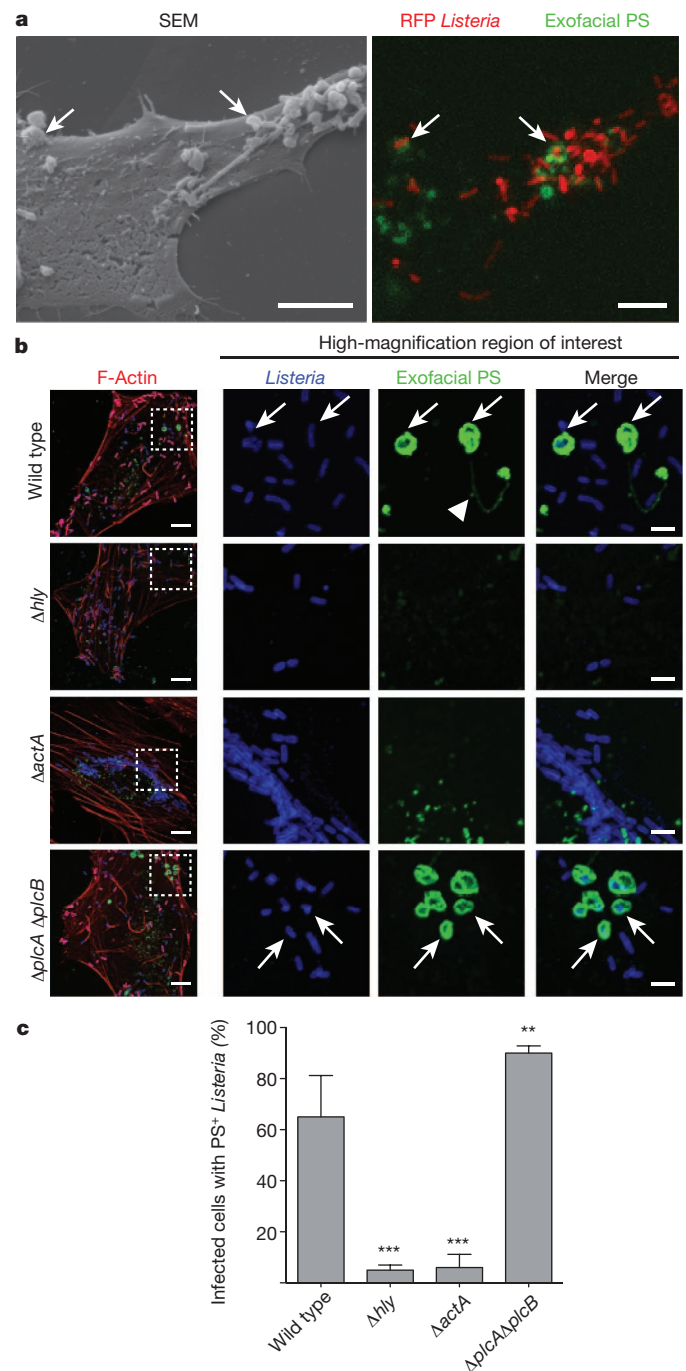


Figure 2 | Formation of PS⁺ structures during *L. monocytogenes* infection. **a**, HeLa cells were infected at an MOI of 100 with wild-type *L. monocytogenes* expressing RFP (red) for 6 h and then incubated with a probe for exofacial PS (annexin A5-Alexa 488; green). Cells were then fixed and analysed by fluorescence microscopy to identify PS⁺ structures and *L. monocytogenes* structures (left). Images are representative of two independent experiments. Scale bars, 5 μm . **b**, Confocal images of cells infected with the indicated strain for 6 h. Boxed areas (left) are low-magnification images (a three-colour merge, includes phalloidin stain for F-actin); they are enlarged on the right. Arrows indicate *L. monocytogenes* bacteria that co-localize with PS, and arrowhead indicates PS⁺ filaments. Images are representative of three independent experiments. Scale bars, 10 μm for low-magnification merge, and 3 μm for enlarged panels. **c**, Cells containing PS⁺ bacteria enumerated for **b**. Averages \pm s.d. for three independent experiments are shown ($n = 100$). P values were calculated using one-way ANOVA. ** $P < 0.01$, *** $P < 0.001$.

Second, $\Delta actA$ mutants did not form these structures (Fig. 2b, c). Third, we observed that the kinetics of formation of PS⁺ structures were similar to the kinetics of protrusion formation (marked by ezrin⁺; Fig. 3a). Live cell imaging revealed that bacteria-containing protrusions (marked by LifeAct-RFP) rapidly recruited the exofacial PS probe, before rounding of the PS⁺ protrusion into vesicles on the cell surface (Fig. 3b and Supplementary Video 1). Recruitment of the exofacial PS probe was limited to protrusions and vesicles, indicating that loss of membrane asymmetry was not the result of global cellular membrane damage. Annexin A2 was localized to PS⁺ bacteria (Extended Data Fig. 1c), suggesting that membrane repair pathways serve to locally restrict membrane damage.

Streptolysin O, a pore-forming toxin from group A *Streptococcus*, has been shown to induce membrane blebbing and release of plasma-membrane-derived vesicles^{18,19}. Our studies indicated that *L. monocytogenes* similarly induces the release of membrane vesicles through LLO. PS⁺ structures containing bacteria were observed on coverslips with no connections to infected cells (Extended Data Fig. 7a). We also isolated these vesicles by centrifugation of the medium (Extended Data Fig. 7b). Our findings suggest that bacteria-containing protrusions are released by infected host cells, giving rise to PS⁺ vesicles in the medium and that, in some cases, these vesicles remain associated with the cell surface of infected cells. Deletion of both PLCs caused an increase in the number of PS⁺ structures associated with the surface of infected cells (Fig. 2b, c). Thus, PLCs may promote bacterial escape from PS⁺ structures after their release from infected cells, which would be consistent with the role of PLCs in escape from vacuoles in secondarily infected cells⁴.

Efferocytosis receptors mediate uptake of PS⁺ cells and cellular debris¹. In macrophages, TIM-4 has a key role in efferocytosis^{20,21}. Because macrophages are a major target of *L. monocytogenes* during systemic infection, we examined the role of TIM-4 in cell-to-cell spread by these bacteria using an infection focus assay. The foci of infection in BMDMs from *Timd4*^{-/-} mice contained fewer infected cells compared with control BMDMs from C57BL/6 mice (Fig. 4a, b). Using live cell imaging, we observed that infection foci were smaller in BMDM cultures from *Timd4*^{-/-} mice from 8 to 18 h post-infection (Fig. 4c, d). Bacterial replication within primarily infected *Timd4*^{-/-} BMDMs at low cell density (minimizing cell-to-cell spread) was comparable with that of control BMDMs (Extended Data Fig. 8). This indicated that deficient bacterial spread in *Timd4*^{-/-} BMDM cultures was not due to impaired bacterial replication in the cytosol of these cells. Blocking antibodies that targeted either TIM-4 or PS impaired bacterial spread in cultures of control but

not *Timd4*^{-/-} BMDMs (Fig. 4b), indicating that TIM-4 promotes bacterial cell-to-cell spread through its ability to bind PS⁺ structures.

TIM-4 was previously shown to suppress pro-inflammatory cytokine production *in vivo*²². Therefore, we considered the possibility that TIM-4 deficiency impairs cell-to-cell spread by *L. monocytogenes* indirectly, through enhanced pro-inflammatory cytokine production. However, in our *in vitro* experiments, pro-inflammatory cytokine production was typically not observed until 18 h post-infection (after bacterial spread is initiated) and shared the same profile of expression in both control and *Timd4*^{-/-} BMDM cultures (Extended Data Fig. 9a). This suggested that TIM-4 acts in a cell-autonomous manner to promote cell-to-cell spread by *L. monocytogenes*.

To test this hypothesis, we used a cell-to-cell spread assay that measures the direct transmission of bacteria from primary infected cells to secondary cells (labelled with CellTrackerBlue). Primary cells were infected with a bacterial mutant lacking both PLCs ($\Delta plcA\Delta plcB$), as this prevents bacterial escape from spreading vacuoles and avoids the confounding effect of rapid replication in the cytosol of secondary cells. Bacterial spread to secondary BMDMs from *Timd4*^{-/-} mice was reduced compared with control mice (Fig. 4e). Microscopic analysis revealed higher numbers of actin⁺ protrusions associated with secondary BMDMs from control mice compared with *Timd4*^{-/-} mice (Fig. 4f, g). This suggests that TIM-4 may promote the association and/or stability of protrusions as they project from infected cells into adjacent cells. Together, our findings indicate a specific role for TIM-4 in mediating cell-to-cell spread by *L. monocytogenes*.

We observed decreased numbers of bacteria in the liver and spleen of *Timd4*^{-/-} mice compared with control mice (Fig. 4h). Similar numbers of $\Delta actA$ mutant bacteria were observed in the liver of control and *Timd4*^{-/-} mice (Fig. 4i), indicating that TIM-4 is required for pathogenesis of motile bacteria in this organ and consistent with its role in promoting bacterial cell-to-cell spread. However, we observed a decrease in survival of $\Delta actA$ mutant bacteria in spleens of infected *Timd4*^{-/-} mice. We also observed elevated basal levels of pro-inflammatory cytokines in the spleens (and to a lesser extent the livers) of *Timd4*^{-/-} mice before infection (Extended Data Fig. 9b). Together, our findings suggest that *L. monocytogenes* exploits TIM-4 during systemic infection of mice by two mechanisms; directly, through uptake of PS⁺ bacteria, and indirectly, through the ability of TIM-4 to suppress pro-inflammatory cytokine production.

It has previously been demonstrated that TIM-4 mediates the clearance of apoptotic macrophages bearing *Mycobacterium tuberculosis*²³.

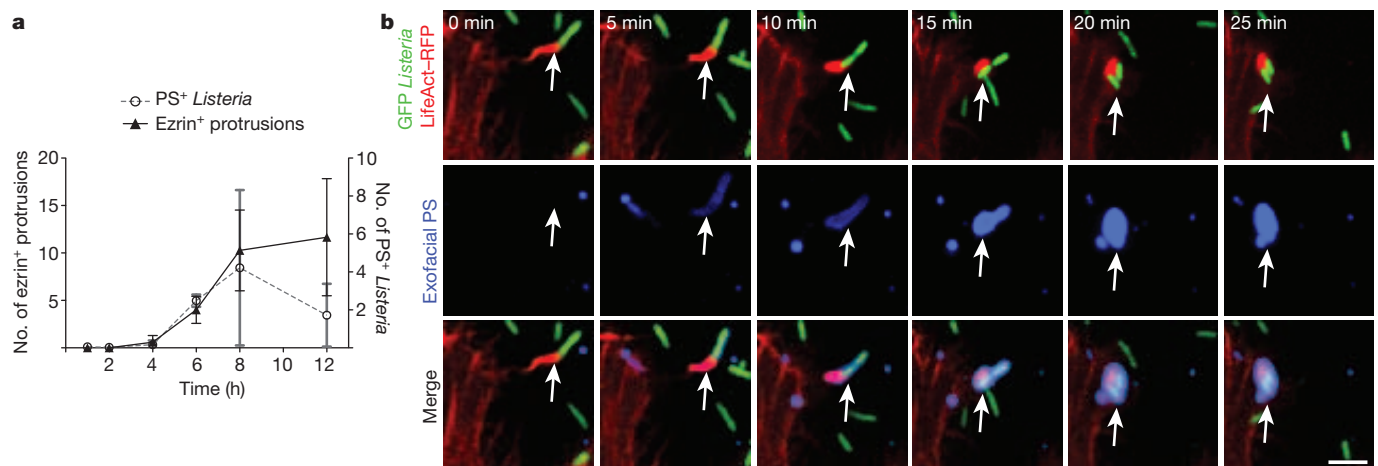
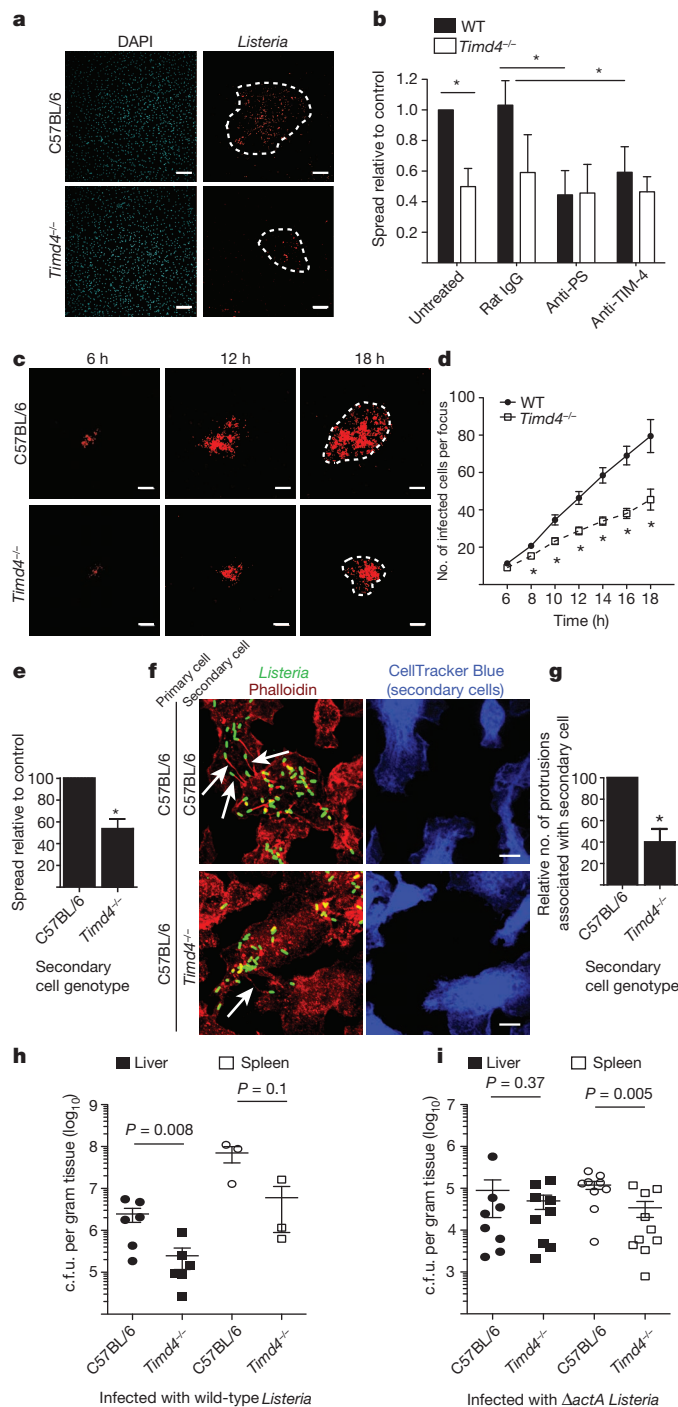


Figure 3 | Protrusions give rise to PS⁺ vesicles containing *L. monocytogenes*. **a**, PS⁺ bacteria and ezrin⁺ protrusions were enumerated in HeLa cells infected at an MOI of 100 with wild-type *L. monocytogenes* for the indicated time. Averages \pm s.d. for three independent experiments are shown ($n = 100$). **b**, Cells were transfected with LifeAct-RFP (F-actin probe)

and then infected with wild-type *L. monocytogenes* expressing GFP for 6 h. Live infected cells were analysed by spinning-disk confocal microscopy with annexin A5-Alexa 647 in the medium to label exofacial PS (green). Successive frames are shown. Arrows indicates protrusion that acquires PS. Images are representative of four independent experiments. Scale bar, 10 μ m.



This finding indicates an important role for efferocytosis in innate immunity. By contrast, our findings suggest that efferocytosis can be exploited by *L. monocytogenes* to promote cell-to-cell spread. Other mechanisms also contribute to *L. monocytogenes* dissemination in a host^{24–26}. Viruses are known to exploit efferocytosis by incorporating PS into their viral membrane^{27,28}. *L. monocytogenes* does not incorporate PS into its cellular wall but rather exploits host membrane repair mechanisms to promote release of protrusion-associated membrane, giving rise to a PS⁺ ‘cloak’ that can promote efferocytosis (see model in Extended Data Fig. 10). Similar forms of non-lytic ejection of bacteria from host cells have been described for *Mycobacterium marinum*²⁹ and *Chlamydia trachomatis*³⁰, indicating that these (and probably other) pathogens may also exploit efferocytosis during infection. In summary, our study identifies a novel strategy used by pathogenic bacteria to promote their

Figure 4 | TIM-4 promotes *L. monocytogenes* cell-to-cell spread in macrophages and growth in mice. **a**, Infection focus assay for measuring cell-to-cell spread. BMDMs were infected with *L. monocytogenes* at an MOI of 0.01. Images are representative of three independent experiments. Dotted lines delineate infection foci. Scale bars, 100 μ m. **b**, Infection foci from **a** were enumerated by fluorescence microscopy. A total of 100 infection foci were analysed for each genotype/strain. Where indicated, cells were treated with anti-PS, anti-TIM-4 or non-specific rat IgG (control) during infection. WT, wild type. Averages \pm s.d. for three independent experiments are shown. *P* values were calculated using two-way ANOVA with Bonferroni post-test. **c**, Monolayers of control or Timd4^{-/-} BMDMs were infected at an MOI of 0.01 with *L. monocytogenes* expressing RFP and examined by live cell imaging. Images are representative of two independent experiments. Dotted lines delineate infection foci. Scale bars, 100 μ m. **d**, The number of infected cells per focus from **c**. Averages \pm s.d. for two independent experiments are shown (*n* = 10). *P* values were calculated using two-tailed Student’s *t*-test. **e**, Cell-to-cell spread assay. Averages \pm s.d. for three independent experiments are shown (*n* = 100). *P* values were calculated using two-tailed Student’s *t*-test. **f**, Images from **e**. Arrows indicate actin⁺ protrusions associated with secondary cells. Images are representative of four independent experiments (*n* = 100). Scale bars, 10 μ m. **g**, Actin⁺ protrusions associated with secondary cells were enumerated. Averages \pm s.d. for four independent experiments are shown. **h**, **i**, Mice were infected with 5×10^4 colony-forming units (c.f.u.) wild-type (**h**) or 1×10^7 c.f.u. $\Delta actA$ mutant (**i**) *L. monocytogenes* by tail vein injection. Mice were killed and livers and spleen were harvested at 72 h after infection for quantification of bacterial load (c.f.u. per gram tissue). Data are expressed as mean \pm standard error of the mean. *P* values were calculated using one-tailed Mann–Whitney test.

dissemination in a host and provides a remarkable example of how pathogens can exploit innate immune defences during infection.

METHODS SUMMARY

All *L. monocytogenes* strains were derivatives of 10403S and are listed in Methods. Six-to-eight-week-old C57BL/6 and Timd4^{-/-} mice were used to generate BMDMs or were infected intravenously for systemic infection studies. Cytokine levels were measured using a mouse 7-plex pro-inflammatory kit. *In vitro* infections of confluent cell monolayers were performed at an MOI of 0.1–100, as indicated in figure legends, and were performed in the presence of gentamicin to prevent extracellular growth. Constructs, antibodies, siRNA and other pharmacological reagents are listed in Methods. PI staining was carried out 6 h post-infection in Tyrode’s buffer in the presence or absence of calcium. Annexin A5–Alexa 488 was added to chilled cells for exofacial PS staining. Staining protocols are indicated in Methods. For the infection focus assay, confluent monolayers of BMDMs were infected for 18 h and the size of infection foci was analysed. For the direct cell-to-cell spread assay, primary cells were lifted with CellStripper and overlaid onto fresh cultures of CellTrackerBlue-labelled cells for 90 min.

Unless indicated, cells were fixed with paraformaldehyde. Quantifications were done on a Leica DMIRE2 epifluorescence microscope with $\times 63$ objective. Correlative electron microscopy images were acquired from dehydrated samples gold coated in a Denton Desk II sputter coater and examined under a FEI XL30 SEM. Confocal *z* stacks (2 μ m) were taken with a Quorum spinning disk microscope, analysed with Volocity 6 software and imported into Adobe Illustrator. Nikon Ti-E microscopy was used to acquire images of infection foci at $\times 10$ magnification.

The mean plus s.d. is shown in figures. *P* values were calculated using GraphPad Prism v.4.0a using statistical tests indicated in figure legends. A *P* value less than 0.05 was considered significant.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 22 February 2013; accepted 21 February 2014.

Published online 13 April 2014.

1. Ravichandran, K. S. Beginnings of a good apoptotic meal: the find-me and eat-me signaling pathways. *Immunity* **35**, 445–455 (2011).
2. Mostowy, S. & Cossart, P. Virulence factors that modulate the cell biology of *Listeria* infection and the host response. *Adv. Immunol.* **113**, 19–32 (2012).
3. Robbins, J. R. et al. *Listeria monocytogenes* exploits normal host cell processes to spread from cell to cell. *J. Cell Biol.* **146**, 1333–1350 (1999).
4. Alberti-Segui, C., Goeden, K. R. & Higgins, D. E. Differential function of *Listeria monocytogenes* listeriolysin O and phospholipases C in vacuolar dissolution following cell-to-cell spread. *Cell. Microbiol.* **9**, 179–195 (2007).

5. Gedde, M. M., Higgins, D. E., Tilney, L. G. & Portnoy, D. A. Role of listeriolysin O in cell-to-cell spread of *Listeria monocytogenes*. *Infect. Immun.* **68**, 999–1003 (2000).
6. Schnupf, P. & Portnoy, D. A. Listeriolysin O: a phagosome-specific lysin. *Microbes Infect.* **9**, 1176–1187 (2007).
7. Glomski, I. J., Gedde, M. M., Tsang, A. W., Swanson, J. A. & Portnoy, D. A. The *Listeria monocytogenes* hemolysin has an acidic pH optimum to compartmentalize activity and prevent damage to infected host cells. *J. Cell Biol.* **156**, 1029–1038 (2002).
8. Nomura, T. *et al.* Irreversible loss of membrane-binding activity of *Listeria*-derived cytolysins in non-acidic conditions: a distinct difference from allied cytolysins produced by other Gram-positive bacteria. *Microbiology* **153**, 2250–2258 (2007).
9. Schnupf, P., Portnoy, D. A. & Decatur, A. L. Phosphorylation, ubiquitination and degradation of listeriolysin O in mammalian cells: role of the PEST-like sequence. *Cell. Microbiol.* **8**, 353–364 (2006).
10. Hamon, M. A., Ribet, D., Stavru, F. & Cossart, P. Listeriolysin O: the Swiss army knife of *Listeria*. *Trends Microbiol.* **20**, 360–368 (2012).
11. Cassidy, S. K. *et al.* Membrane damage during *Listeria monocytogenes* infection triggers a caspase-7 dependent cytoprotective response. *PLoS Pathog.* **8**, e1002628 (2012).
12. Idone, V., Tam, C. & Andrews, N. W. Two-way traffic on the road to plasma membrane repair. *Trends Cell Biol.* **18**, 552–559 (2008).
13. Gründling, A., Gonzalez, M. D. & Higgins, D. E. Requirement of the *Listeria monocytogenes* broad-range phospholipase PC-PLC during infection of human epithelial cells. *J. Bacteriol.* **185**, 6295–6307 (2003).
14. Draeger, A., Monastyrskaya, K. & Babiychuk, E. B. Plasma membrane repair and cellular damage control: the annexin survival kit. *Biochem. Pharmacol.* **81**, 703–712 (2011).
15. Fadeel, B. & Xue, D. The ins and outs of phospholipid asymmetry in the plasma membrane: roles in health and disease. *Crit. Rev. Biochem. Mol. Biol.* **44**, 264–277 (2009).
16. Waite, J. C. *et al.* Dynamic imaging of the effector immune response to *Listeria* infection *in vivo*. *PLoS Pathog.* **7**, e1001326 (2011).
17. Pust, S., Morrison, H., Wehland, J., Sechi, A. S. & Herrlich, P. *Listeria monocytogenes* exploits ERM protein functions to efficiently spread from cell to cell. *EMBO J.* **24**, 1287–1300 (2005).
18. Babiychuk, E. B., Monastyrskaya, K., Potez, S. & Draeger, A. Blebbing confers resistance against cell lysis. *Cell Death Differ.* **18**, 80–89 (2011).
19. Keyel, P. A. *et al.* Streptolysin O clearance through sequestration into blebs that bud passively from the plasma membrane. *J. Cell Sci.* **124**, 2414–2423 (2011).
20. Feng, D. *et al.* Cellular internalization of exosomes occurs through phagocytosis. *Traffic* **11**, 675–687 (2010).
21. Miyanishi, M. *et al.* Identification of Tim4 as a phosphatidylserine receptor. *Nature* **450**, 435–439 (2007).
22. Rodriguez-Manzanet, R. *et al.* T and B cell hyperactivity and autoimmunity associated with niche-specific defects in apoptotic body clearance in TIM-4-deficient mice. *Proc. Natl Acad. Sci. USA* **107**, 8706–8711 (2010).
23. Martin, C. J. *et al.* Efferocytosis is an innate antibacterial mechanism. *Cell Host Microbe* **12**, 289–300 (2012).
24. Appelberg, R. & Leal, I. S. Mutants of *Listeria monocytogenes* defective in *in vitro* invasion and cell-to-cell spreading still invade and proliferate in hepatocytes of neutropenic mice. *Infect. Immun.* **68**, 912–914 (2000).
25. Drevets, D. A. Dissemination of *Listeria monocytogenes* by infected phagocytes. *Infect. Immun.* **67**, 3512–3517 (1999).
26. Friedrich, N., Hagedorn, M., Soldati-Favre, D. & Soldati, T. Prison break: pathogens' strategies to egress from host cells. *Microbiol. Mol. Biol. Rev.* **76**, 707–720 (2012).
27. Meertens, L. *et al.* The TIM and TAM families of phosphatidylserine receptors mediate dengue virus entry. *Cell Host Microbe* **12**, 544–557 (2012).
28. Mercer, J. & Helenius, A. Vaccinia virus uses macropinocytosis and apoptotic mimicry to enter host cells. *Science* **320**, 531–535 (2008).
29. Hagedorn, M., Rohde, K. H., Russell, D. G. & Soldati, T. Infection by tubercular mycobacteria is spread by nonlytic ejection from their amoeba hosts. *Science* **323**, 1729–1733 (2009).
30. Hybishe, K. & Stephens, R. S. Mechanisms of host cell exit by the intracellular bacterium *Chlamydia*. *Proc. Natl Acad. Sci. USA* **104**, 11430–11435 (2007).

Supplementary Information is available in the online version of the paper.

Acknowledgements We are grateful to S. Gray-Owen, S. Grinstein and D. Portnoy for providing reagents and advice and to D. Holmyard for help with electron microscopy. J.H.B. holds the Pitblado Chair in Cell Biology. Infrastructure for the Brumell laboratory was provided by a Leader's Opportunity Fund grant from the Canadian Foundation for Innovation and the Ontario Innovation Trust. R.F. was supported by a postdoctoral fellowship from the Canadian Institutes of Health Research in partnership with the Canadian Association of Gastroenterology and the Crohn's and Colitis Foundation of Canada. S.O. was supported by a postdoctoral fellowship from the Research Training Committee at the Hospital for Sick Children. This work was supported by an operating grant from The Arthritis Society of Canada (#RG11/013) to J.H.B. and a US Public Health Service grant (AI053669) from the National Institutes of Health to D.E.H.

Author Contributions J.H.B., M.A.C., S.O. and D.E.H. designed the experiments and wrote the paper. M.A.C., R.F., J.M.v.R., V.C. and S.O. performed the experiments. A.M.M. and V.K.K. contributed reagents and consultations.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.H.B. (john.brumell@sickkids.ca).

METHODS

Bacterial strains. *L. monocytogenes* were grown in brain-heart infusion (BHI) broth and the following strains were used: 10403S (wild type)³¹, DP-L2161 (Δhly)³², DP-L4818 ($\Delta hly + hly$)³³, DP-L3078 ($\Delta actA$)³⁴, and DP-L1936 ($\Delta plcA \Delta plcB$)³⁵. Wild-type *L. monocytogenes* 10403S expressing TagRFP under the *actA* promoter (DP-L5538)¹⁶ or expressing GFP under the *hly* promoter (DP-L1039) were previously described³⁶.

Antibodies, constructs and reagents. Primary antibodies used were rabbit anti-annexin A2 (#610068 from BD Biosciences), rabbit anti-GFP (#A11120 from Invitrogen), rabbit anti-caspase 7 (#12827 from Cell Signaling), rabbit anti-*Listeria* (#B223021 from BD Biosciences), rabbit anti-*L. monocytogenes* antibody (gift from P. Cossart), mouse anti-phosphatidylserine (#18005 from Abcam) and mouse anti-ezrin (#35-7300 from Invitrogen). Rat anti-TIM-4 blocking antibody was previously described³⁷. Alexa Fluor-568 Phalloidin, annexin A5–Alexa Fluor-488 and -647 conjugates and all fluorescent secondary antibodies (Alexa Fluor conjugates) were from Invitrogen. DAPI (#D1306 from Invitrogen) was used at 1:2,500 dilution to stain the nuclei where indicated. Cytochalasin D (#2502555; 10 μ M final) and Latrunculin B (#428020; 10 μ M final) were from Calbiochem. For transfection of HeLa cells, Xtreme Gene 9 (Roche) transfection reagent was used as per manufacturer's protocols. LifeAct–mRFP³⁸ was a gift from R. Truant. The following siRNAs were from Sigma: annexin A1 (#00157996), annexin A2 (#00246294), annexin A6 (#00063383), caspase 7 (#00128361).

Cell culture and macrophage generation. HeLa and Henle-407 cells were cultured in DMEM (Hyclone) supplemented with 10% heat-inactivated FBS (Wisent) without antibiotics at 37 °C and 5% CO₂.

All experimental protocols involving mice were approved by the Animal Care Committee of The Hospital for Sick Children. Mice were euthanized by CO₂ inhalation. Mouse BMDMs were obtained from the dissected femurs and tibias of 5–8-week-old mice. Cells were washed with growth medium and plated on 70 cm² Petri dishes. Medium was replaced every 2 days and after 7–9 days cells were used for experiments. Cells were maintained in high-glucose RPMI-1640 medium (Wisent #350-025-CL) containing 10% heat-inactivated FBS (Wisent), 1% sodium pyruvate (Wisent), 1% non-essential amino acids (Wisent), 0.5% 2-mercaptoethanol (Gibco), 1% penicillin and streptomycin (Invitrogen), and 10% L929 medium. L929 medium was generated by growing a confluent layer of L929 cells in 175 cm² flasks in DMEM supplemented with 10% heat-inactivated FBS. When cells reached confluency, growth medium was replaced by DMEM alone. After 7 days, supernatant was collected, centrifuged, filtered, and stored at –20 °C.

PI assay for membrane damage. HeLa cells were plated at 5×10^4 cells per well in 24-well tissue culture plates ~18 h before infection. For siRNA treatments, the cells were plated at 2.5×10^4 cells per well. siRNAs were applied after 18 h and the media was changed 24 h later. Infections were carried out 24 h later. All strains of *L. monocytogenes* were infected at an MOI of 100. Bacteria were spun onto cells by centrifugation at 1,500 r.p.m. for 5 min. After 60 min of invasion at 37 °C, cells were washed three times with PBS with calcium and magnesium (PBS+, Wisent #311-420-CL) followed by the addition of growth medium containing 50 μ M gentamicin (Wisent #400-130-IG). At 6 h post-infection, cells were washed two times with PBS without calcium and magnesium (PBS–, Wisent #311-010-CL) and replaced with Tyrode's Buffer (10 mM HEPES, 10 mM glucose, 5 mM potassium chloride, 140 mM sodium chloride, 1 mM EGTA, 1 mM magnesium chloride, 2 mM calcium chloride, pH 7.4) containing 0.5 mg ml^{–1} PI (Sigma #P4170). For Tyrode's Buffer lacking calcium the 2 mM calcium chloride was replaced with 2 mM magnesium chloride. Cells were incubated in Tyrode's Buffer with PI for the indicated times and then fixed with 4% PFA (EM Sciences #15710). Samples were mounted on slides using fluorescence mounting medium (Dako). Quantifications were done using a Leica DMIRE2 epifluorescence microscope equipped with a $\times 40$ objective.

Annexin A5–Alexa 488 staining of exofacial PS. HeLa cells were plated at 5×10^4 cells per well in 24-well tissue culture plates ~18 h before infection. All strains of *L. monocytogenes* were infected at an MOI of 100 in DMEM. Bacteria were spun onto cells by centrifugation at 1,500 r.p.m. for 5 min. After 60 min of invasion at 37 °C, cells were washed three times with PBS+ followed by the addition of growth media containing 50 μ M gentamicin (Wisent #400-130-IG). After indicated times of infection, cells were cooled on ice and washed twice with chilled PBS+. Annexin A5–Alexa Fluor-488 conjugate (Invitrogen) was diluted to 1% (v/v) in chilled PBS with calcium and magnesium and added onto the coverslips for 10 min on ice. Cells were washed twice with chilled PBS+ and fixed with 2.5% paraformaldehyde (PFA) (EM Sciences #15710) for 30 min at 37 °C. Samples were mounted on slides using fluorescence mounting medium (Dako).

Immunofluorescence. Immunostaining was conducted as previously described³⁹. Briefly, cells were permeabilized and blocked in PBS containing 0.2% saponin (Calbiochem) and 10% normal goat serum (SS-PBS) for 30 min. Subsequently cells were incubated for 1 h with primary antibodies in SS-PBS. Cells were washed three times with PBS and incubated with secondary Alexa Fluor conjugated antibodies

for 1 h. Cells were washed three times with PBS, mounted in fluorescence mounting medium (Dako), and analysed using a Leica DMIRE2 epifluorescence microscope. Confocal pictures were taken using a Quorum spinning disk microscope (Leica DMIRE2 inverted fluorescence microscope equipped with a Hamamatsu back-thinned electron multiplying charge-coupled device camera, spinning disc head, and Velocity 4 software (Improvision)). Confocal images were imported into Adobe Photoshop and assembled in Adobe Illustrator for labelling.

Correlative light-scanning electron microscopy. HeLa cells were plated at 5×10^5 cells per well in 6-well tissue culture plates with etched grid coverslips with imprinted numbers (BELLCO Biotechnology) 24 h before infection. Cells were infected with wild-type *L. monocytogenes* expressing RFP (DP-L5538) at an MOI of 100 in DMEM. Bacteria were spun onto cells by centrifugation at 1,500 r.p.m. for 5 min. After 60 min of invasion at 37 °C, cells were washed three times with PBS+ followed by the addition of growth media containing 50 μ M gentamicin (Wisent #400-130-IG). At 6 h post-infection, cells were cooled on ice and washed twice with chilled PBS+. Annexin A5–Alexa Fluor 488 conjugate (Invitrogen) was diluted to 1% (v/v) in chilled PBS with calcium and magnesium and added onto the coverslips for 10 min on ice. Cells were washed twice with chilled PBS+ and fixed with 2.5% PFA (EM Sciences #15710) for 30 min at 37 °C. Coverslips were imaged by fluorescence microscopy in PBS+. Subsequently, samples were fixed in 2% glutaraldehyde in cacodylate buffer, rinsed in buffer and dehydrated in a graded ethanol series. The samples were critical point dried in a Bal-tec CPD030 critical point dryer, mounted on aluminium stubs, gold coated in a Denton Desk II sputter coater and examined in an FEI XL30 SEM.

Live cell imaging of PS⁺ structure formation. HeLa cells were plated at 2×10^5 cells per well in 6-well tissue culture plates with glass coverslips 48 h before infection. At 24 h before infection, cells were transfected with LifeAct–RFP. Cells were then infected with wild-type *L. monocytogenes* expressing GFP (DP-L1039) at an MOI of 100 in DMEM. After 60 min of invasion at 37 °C, cells were washed three times with PBS+ followed by the addition of growth media containing 50 μ M gentamicin (Wisent #400-130-IG). At 6 h post-infection, coverslips were washed with PBS+ and transferred into microscope chambers and RPMI-1640 with 10% heat-inactivated FBS (Wisent), 50 μ M gentamicin, 2.5 mM CaCl₂ and 2% (v/v) annexin A5–Alexa Fluor 647 conjugate (Invitrogen). HeLa cells were maintained at 37 °C during imaging. A Leica DMIRE2 inverted fluorescence microscope equipped with a Hamamatsu Back-Thinned EM-CCD camera and spinning disk confocal scan head with a $\times 63$ objective and LSM 510 software was used. Velocity software (Improvision) was used to acquire images.

Animals. *Timd4*^{–/–} mice (on a C57BL/6 background) were previously characterized²² and bred in house at the Hospital for Sick Children Animal Care Facility. C57BL/6 mice, originally from The Jackson Laboratory, were also bred in house and used as controls. All experiments were performed with 7–9-week-old female mice that were maintained on a 12 h light–dark cycle, with food and water available *ad libitum*. Sample size was chosen based on litter size and where possible five mice were used per experiment per group and per bacterial strain. Experiments were not blinded and mice were not randomized. All experiments described in this study were carried out in accordance with the Guide for the Humane Use and Care of Laboratory Animals and were approved by the Hospital for Sick Children's Animal Care Committee.

Infection focus assay. After 7–10 days of differentiation, BMDMs were washed twice and detached with ice-cold Versene buffer (0.8 mM EDTA, 1 mM glucose in PBS–) for 20 min at 4 °C. BMDMs were then seeded onto coverslips in 24-well tissue culture plates at 8.0×10^5 cells per well to generate a monolayer. After 18 h, the monolayer was infected with wild-type *L. monocytogenes* at an MOI of 0.01 in RPMI-1640. At 30 min post-infection, cells were washed three times with PBS and cultured in RPMI-1640 medium containing 10% FBS. At 60 min post-infection, cells were washed three times with PBS and RPMI-1640 containing 10% FBS, 10% L929 conditioned medium, and 50 μ M gentamicin was added to the cultures. At 18 h post-infection, cells were fixed with 2.5% PFA for 30 min at 37 °C and prepared for fluorescence microscopy. The number of infected cells per focus of infection was quantified using epifluorescence microscopy. Images for analysis were taken with a Hamamatsu Orca R2 camera and Nikon Ti-E microscope with $\times 10$ objective and Velocity software (Improvision).

Live cell imaging of *L. monocytogenes* cell-to-cell spread. BMDMs were plated as a monolayer at 4×10^6 cells per well onto coverslips in a 6-well tissue culture plate. After 24 h, the monolayer was infected with *L. monocytogenes* expressing RFP (DP-L5538) at an MOI of 0.01 for 1 h. Cells were washed three times with PBS and incubated at 37 °C in RPMI containing 10% FBS, 10% L-929 conditioned medium and 50 μ M gentamicin. At 6 h post-infection, the coverslips were imaged in 25 mM HEPES buffered RPMI containing 10% FBS and 50 μ M gentamicin using a Quorum spinning disk confocal microscope (Leica DMIRE2 inverted fluorescence microscope equipped with a Hamamatsu back-thinned electron multiplying charge-coupled device camera, Yokogawa spinning disc head, and Velocity

6 software). Coverslips with BMDMs from control and *Timd4*^{-/-} mice were placed side by side on a dual-chamber heated stage at 37 °C. Over the course of 12 h, 36 µm z stacks with a 2 µm step were taken every 15 min at ten foci of infection per coverslip. The channels for DIC and red fluorescence were acquired throughout the experiment. Image analysis was performed to measure the number of infected cells per infection focus during the course of the experiment.

Cell-to-cell spread assay. BMDMs from C57BL/6 mice were seeded at a density of 1×10^6 in 10 cm² Petri dishes 16 h before experiments. Cultures of $\Delta plcA \Delta plcB$ mutant bacteria were used to infect BMDMs at an MOI of 10. At 60 min post-infection, cells were washed three times with PBS and RPMI medium containing 10% FBS and 50 µg ml⁻¹ gentamycin was added to the cultures. At 3 h post-infection, cells were detached using CellStripper (Mediatech) containing 50 µg ml⁻¹ gentamycin and counted. 1×10^5 infected cells were then overlaid onto fresh cultures of CellTracker Blue-labelled BMDMs from C57BL/6 or *Timd4*^{-/-} mice. At 90 min after overlay, cells were fixed with 2.5% PFA for 30 min at 37 °C and prepared for fluorescence microscopy. The uptake of bacteria from infected primary cells (unlabelled) to secondary cells (CellTracker Blue labelled) was determined by microscopic analysis.

Labelling of secondary host cells with CellTracker Blue (Invitrogen) was performed as follows. Secondary cells were plated onto coverslips in 24-well tissue culture plates at 1.5×10^6 cells per well. At 2 h before overlay, serum-free RPMI medium containing 20 µM CellTracker Blue was added to the culture. After 30 min incubation, cells were washed three times with PBS and RPMI medium containing 10% FBS was added.

Macrophage replication assay. BMDMs were plated at 3×10^5 cells per well in 24-well tissue culture plates 24 h before infection. Cells were then infected with *L. monocytogenes* at an MOI of 1 in RPMI-1640. After 30 min of invasion at 37 °C, cells were washed three times with PBS followed by the addition of RPMI-1640. At 60 min post-infection, media was changed, and RPMI-1640 with 10% FBS containing 50 µg ml⁻¹ gentamicin and 10% L929 was added. Cells were then lysed at 2, 4, 6, 8, 10 and 12 h post-infection with 1% Triton X-100 in PBS. Serial dilutions of the lysates were plated on BHI-agar plates and incubated at 37 °C for 16 h for subsequent quantification of intracellular c.f.u.

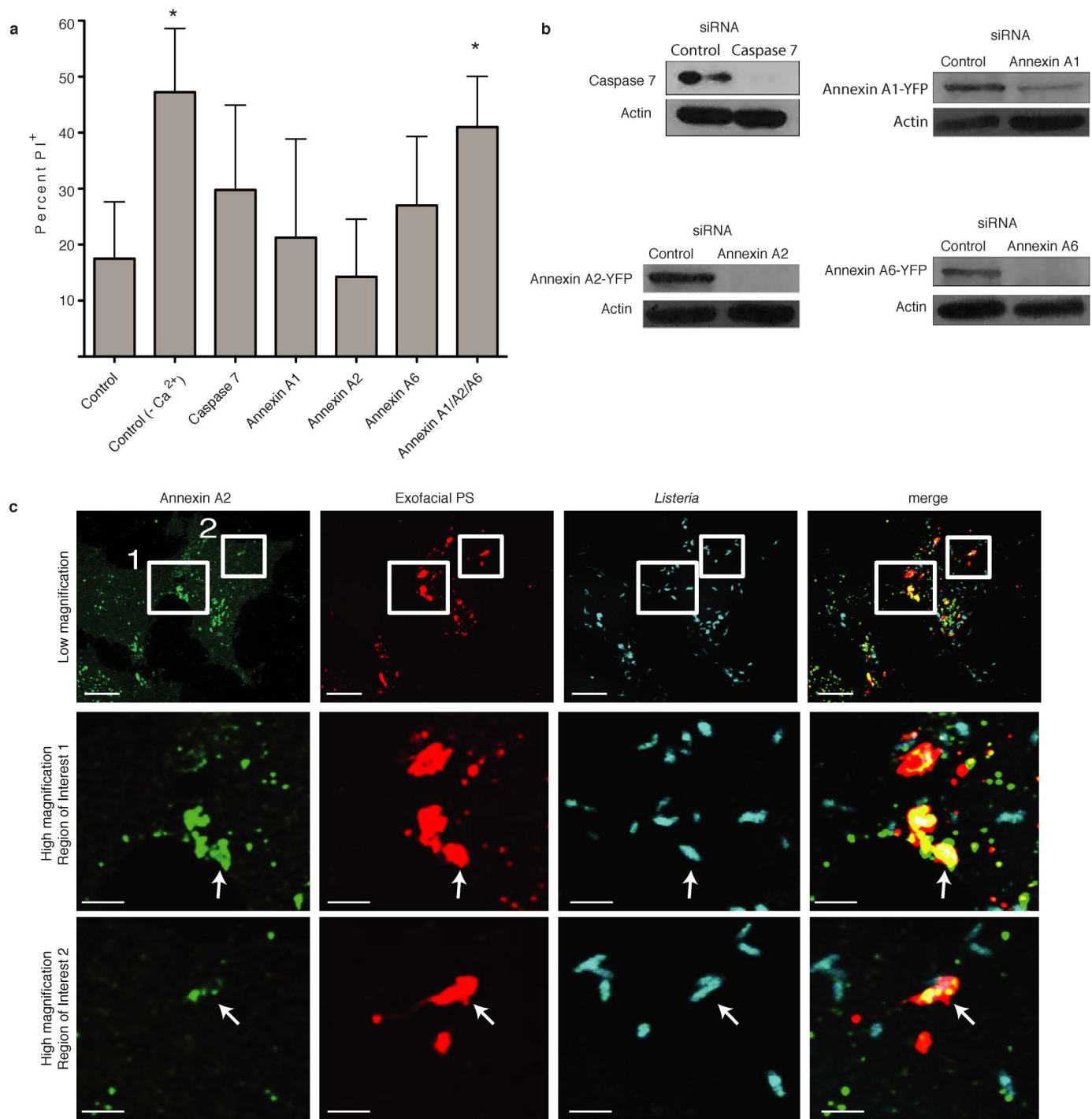
Mouse infection and tissue collection. Mice were infected with 5×10^4 c.f.u. wild-type or 1×10^7 c.f.u. $\Delta actA$ mutant *L. monocytogenes* (in 200 µl of PBS) via

intravenous injection in the lateral tail vein. Liver and spleen were collected at 72 h post-infection. Organs were placed in 1 ml of sterile PBS and homogenized. A dilution series was plated on BHI plates, grown overnight at 37 °C and individual bacterial colonies were enumerated.

For cytokine analysis, cell culture supernatants and tissue homogenates were centrifuged (12,000 r.p.m. for 12 min) and supernatants were collected and stored at -80 °C. Cytokines were measured using the mouse 7-plex pro-inflammatory kit from Meso Scale Discovery, according to the manufacturer's instructions.

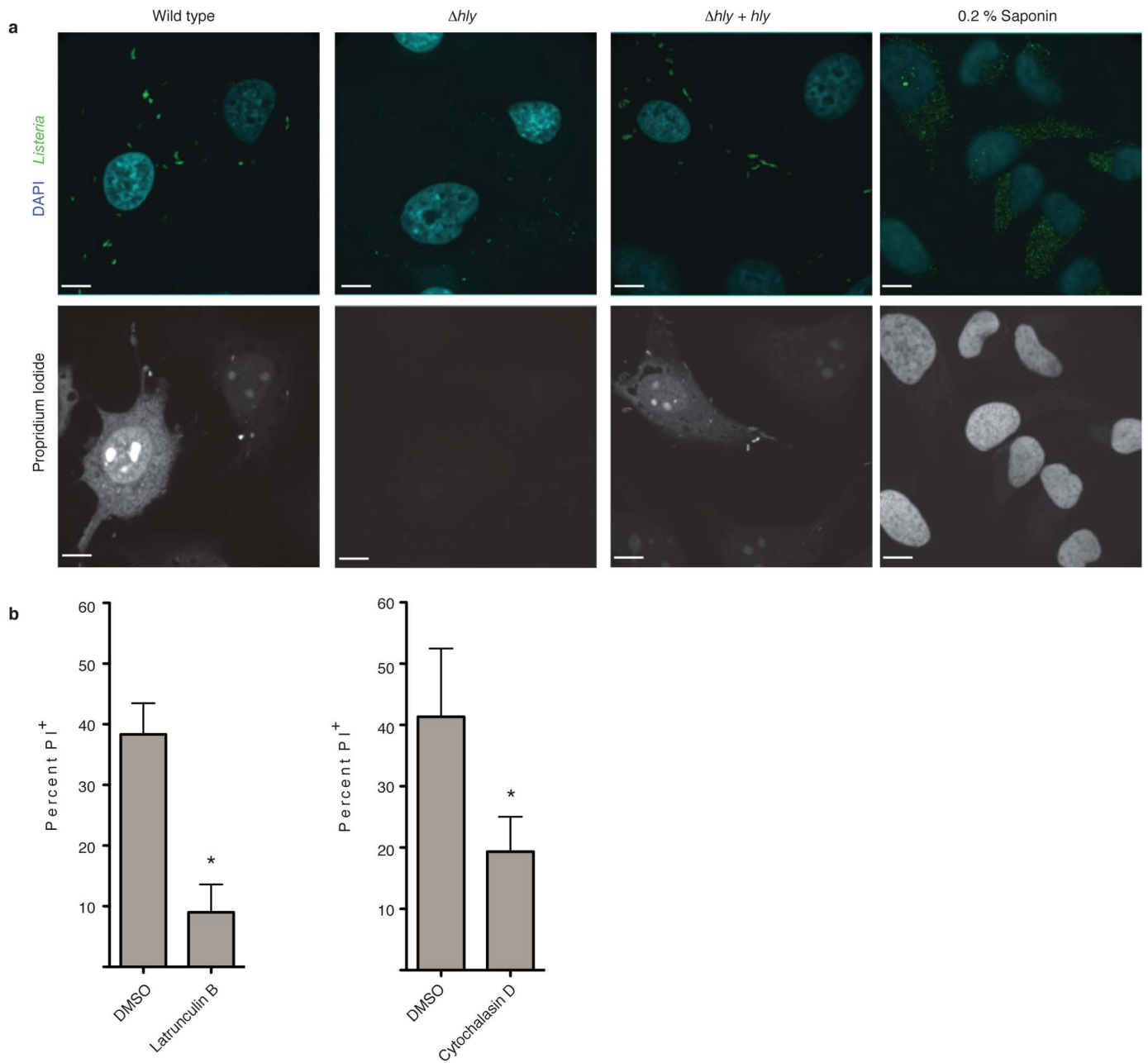
Statistical analysis. Statistical analyses were conducted using GraphPad Prism v.4.0a. The average \pm s.d. is shown in figures, and *P* values were calculated as described in figure legends. A *P* value of less than 0.05 was considered statistically significant.

31. Bishop, D. K. & Hinrichs, D. J. Adoptive transfer of immunity to *Listeria monocytogenes*. The influence of *in vitro* stimulation on lymphocyte subset requirements. *J. Immunol.* **139**, 2005–2009 (1987).
32. Jones, S. & Portnoy, D. A. Characterization of *Listeria monocytogenes* pathogenesis in a strain expressing perfringolysin O in place of listeriolysin O. *Infect. Immun.* **62**, 5608–5613 (1994).
33. Lauer, P., Chow, M. Y., Loessner, M. J., Portnoy, D. A. & Calendar, R. Construction, characterization, and use of two *Listeria monocytogenes* site-specific phage integration vectors. *J. Bacteriol.* **184**, 4177–4186 (2002).
34. Skoble, J., Portnoy, D. A. & Welch, M. D. Three regions within ActA promote Arp2/3 complex-mediated actin nucleation and *Listeria monocytogenes* motility. *J. Cell Biol.* **150**, 527–538 (2000).
35. Smith, G. A. *et al.* The two distinct phospholipases C of *Listeria monocytogenes* have overlapping roles in escape from a vacuole and cell-to-cell spread. *Infect. Immun.* **63**, 4231–4237 (1995).
36. Shen, A. & Higgins, D. E. The 5' untranslated region-mediated enhancement of intracellular listeriolysin O production is required for *Listeria monocytogenes* pathogenicity. *Mol. Microbiol.* **57**, 1460–1473 (2005).
37. Rodriguez-Manzanet, R. *et al.* TIM-4 expressed on APCs induces T cell expansion and survival. *J. Immunol.* **180**, 4706–4713 (2008).
38. Munsie, L. N., Caron, N., Desmond, C. R. & Truant, R. Lifeact cannot visualize some forms of stress-induced twisted F-actin. *Nature Methods* **6**, 317 (2009).
39. Brumell, J. H., Rosenberger, C. M., Gotto, G. T., Marcus, S. L. & Finlay, B. B. SifA permits survival and replication of *Salmonella typhimurium* in murine macrophages. *Cell. Microbiol.* **3**, 75–84 (2001).



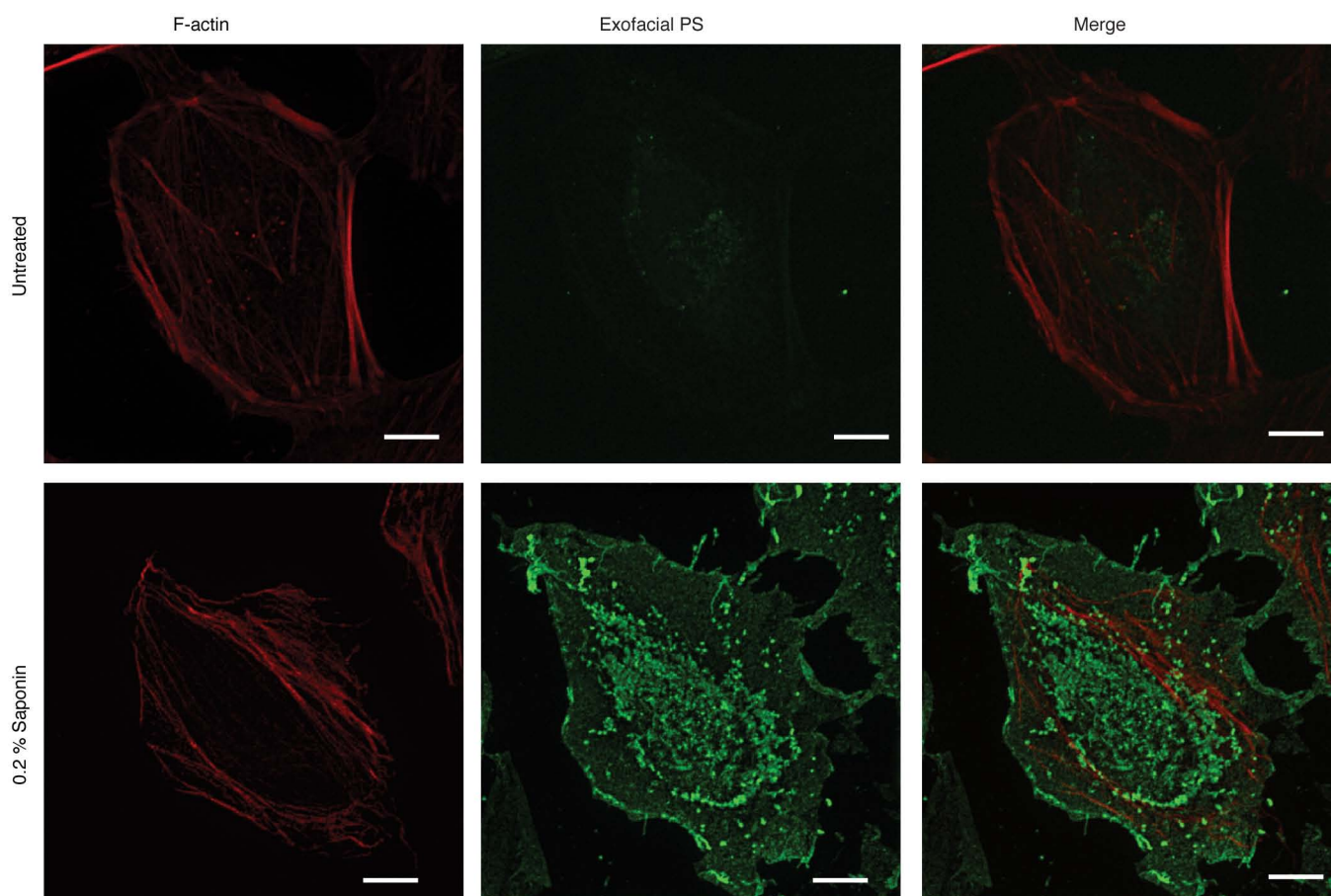
Extended Data Figure 1 | Annexins promote membrane repair during *L. monocytogenes* infection. **a**, HeLa cells were treated with the indicated siRNA for 48 h and then infected with wild-type *L. monocytogenes* at an MOI of 100. At 6 h post-infection, medium was switched to Tyrodes buffer containing 0.5 mg ml⁻¹ PI with or without calcium. Cells were fixed at 60 min after PI addition and then stained for bacteria and DNA (DAPI). The percentage of 100 random infected cells that were PI⁺ cells were enumerated by microscopic analysis. Averages \pm s.d. for three independent experiments are shown.

P values were calculated using one-way ANOVA. **P* < 0.05. **b**, Knockdown of gene expression by siRNA was confirmed by western blotting. Images are representative of two independent experiments. **c**, Recruitment of annexin A2 to PS⁺ structures containing bacteria. Boxes in low-magnification image indicate areas enlarged in bottom panels. Arrows indicate PS⁺ structures that co-localize with annexin A2. Images are representative of three independent experiments. Scale bars, 10 μ m for low-magnification images, 2 μ m for enlarged regions of interest.



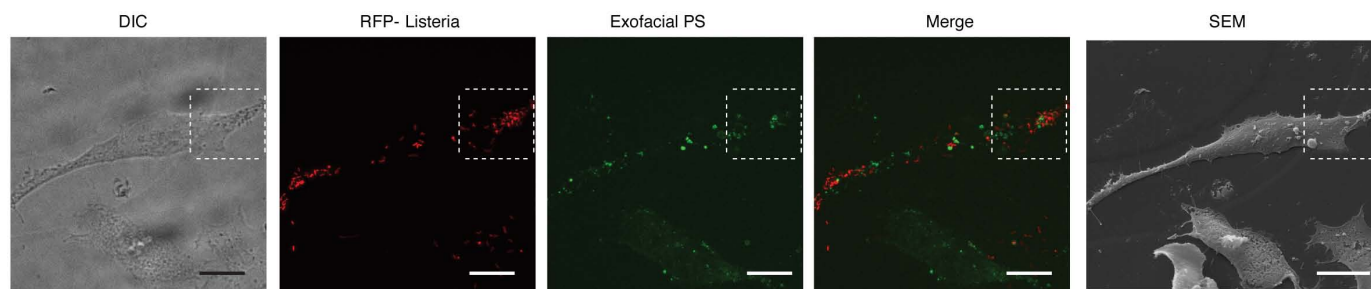
Extended Data Figure 2 | Actin-based motility promotes LLO-mediated membrane damage during *L. monocytogenes* infection. **a**, HeLa cells were infected with the indicated *L. monocytogenes* strain. At 6 h post-infection, medium was switched to Tyrodes buffer containing 0.5 mg ml⁻¹ PI with or without calcium. Cells were fixed at 60 min after PI addition and then stained for bacteria and DNA (DAPI). Confocal images representative of three independent experiments are shown ($n = 100$). PI⁺ cells were enumerated and

results are shown in Fig. 1d. Where indicated, uninfected cells were treated with saponin to permeabilize membranes and allow PI entry, serving as a positive control. Scale bars, 10 μ m. **b**, HeLa cells were infected with wild-type *L. monocytogenes* and subjected to membrane damage assay as in **a** in the presence of either DMSO or the actin cytoskeleton inhibitors latrunculin B or cytochalasin D. Averages \pm s.d. for three independent experiments are shown ($n = 100$). P values were calculated using two-tailed Student's t -test. * $P < 0.05$.



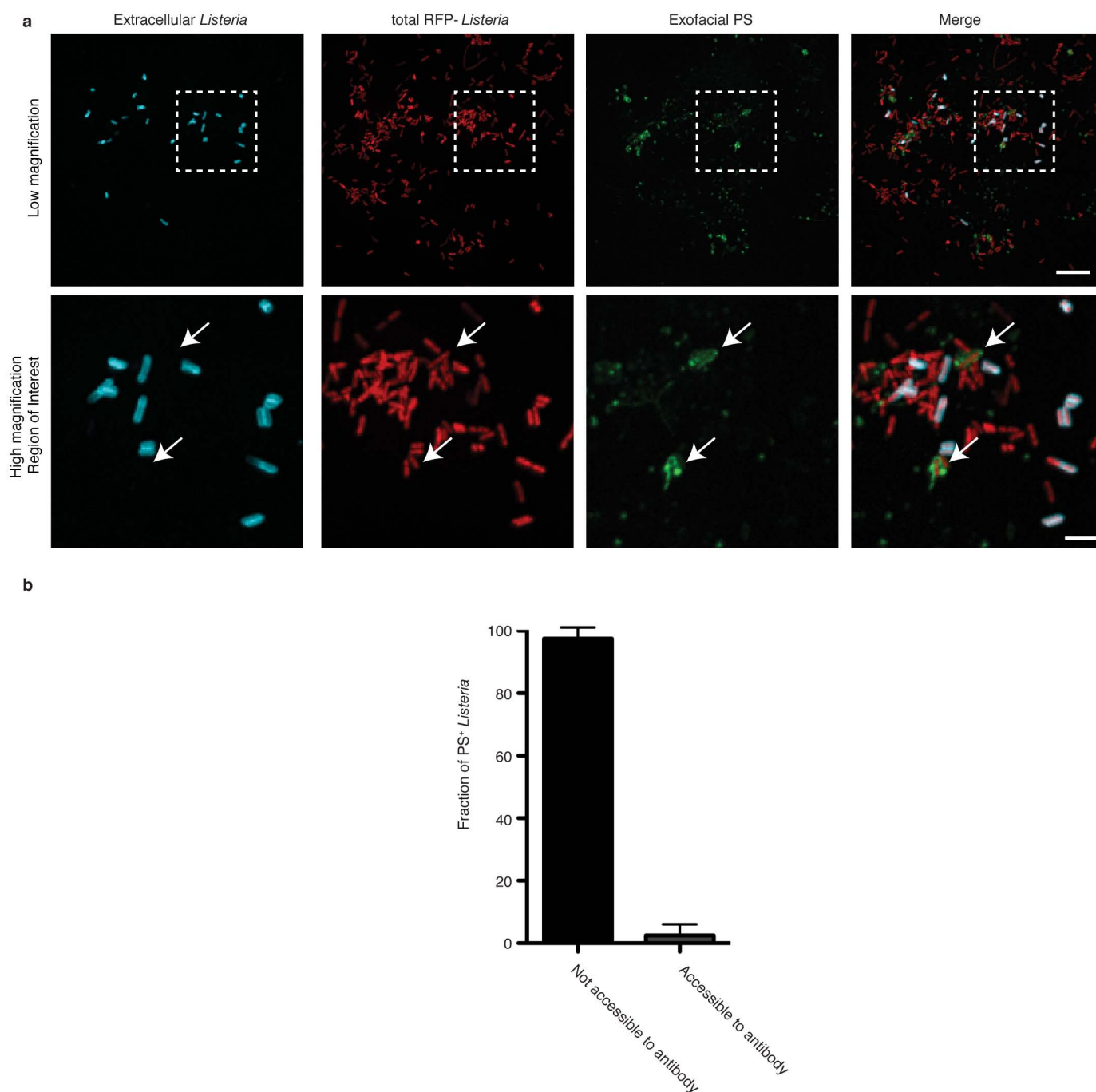
Extended Data Figure 3 | Annexin A5–Alexa 488 as a probe to label PS. Live HeLa cells were cooled on ice and stained with a fluorescent probe (annexin A5–Alexa 488) for 10 min to label exofacial PS. Cells were then fixed and stained with phalloidin Alexa 568 to visualize F-actin. In uninfected control experiments, low amounts of exofacial PS was detected in the membranes of

cells, due to asymmetry of PS distribution in the plasma membrane. By contrast, treatment of cells with the pore-forming surfactant saponin led to robust staining of cells with annexin A5–Alexa 488. Images representative of three independent experiments. Scale bars, 10 μ m.



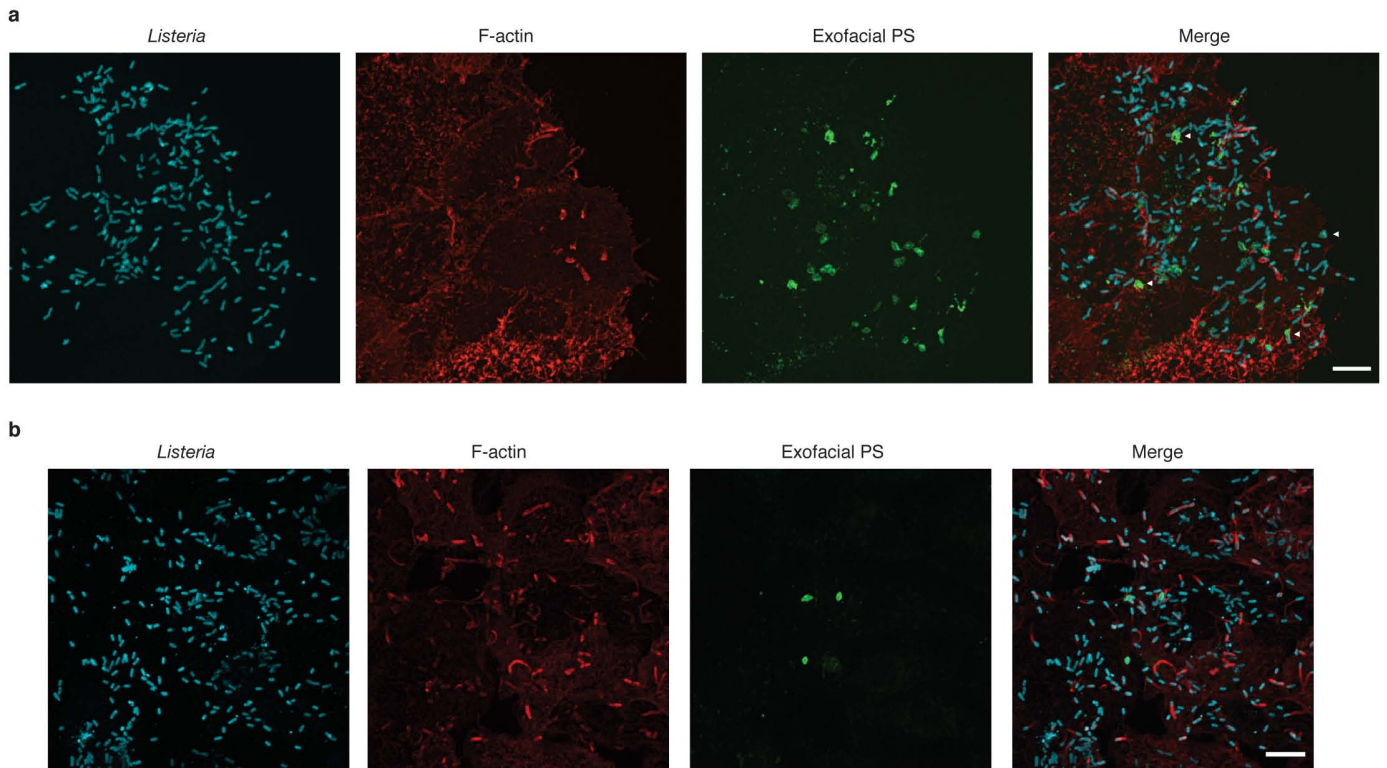
Extended Data Figure 4 | Formation of PS⁺ structures during *L. monocytogenes* infection. Low-magnification images used to generate images shown in Fig. 2a. HeLa cells were infected with wild-type *L. monocytogenes* expressing RFP for 6 h and then cooled on ice and stained with a fluorescent probe (annexin A5–Alexa 488) for 10 min to label exofacial PS. Cells were then fixed and analysed by fluorescence microscopy to

identify PS⁺ structures and bacteria. SEM of the same cell revealed that PS⁺ structures were associated with the dorsal surface of infected cells. Differential interference contrast (DIC) microscopy of cells was also performed to help identify cells for correlative imaging analysis. Images are representative of two independent experiments. Scale bars, 20 μ m.



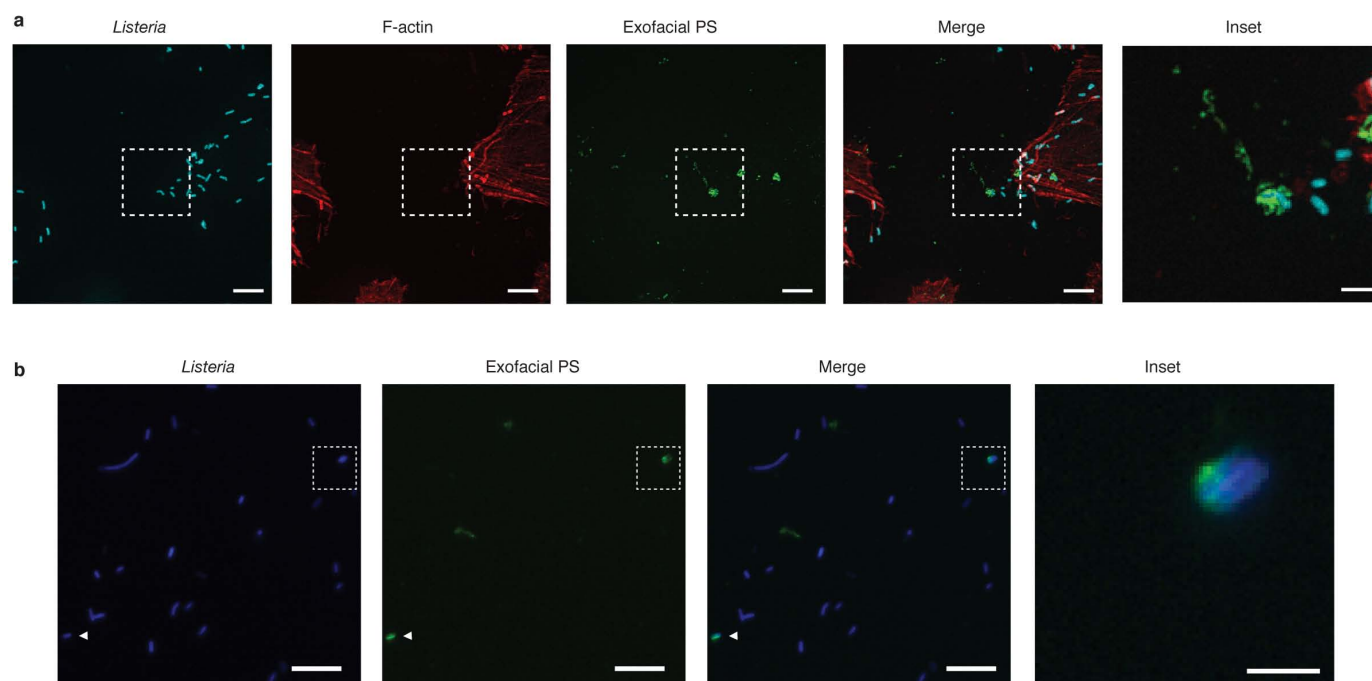
Extended Data Figure 5 | PS⁺ bacteria are present with a host-derived membrane structure. **a**, HeLa cells were infected with wild-type *L. monocytogenes* expressing RFP for 8 h and then labelled with a probe for exofacial PS (annexin A5–Alexa 488). Cells were then rapidly stained with anti-*Listeria* antibodies (5 min) to label extracellular bacteria. Cells were then fixed and analysed by fluorescence microscopy to identify PS⁺ structures and bacteria. Bacteria that co-localize with exofacial PS but are not labelled with anti-*Listeria* antibodies in the extracellular medium are indicated with arrows. Extracellular bacteria do not label with annexin A5–Alexa 488, indicating that

this probe does not bind non-specifically to bacteria. Box in low-magnification image indicates area enlarged in bottom panels. Images are representative of three independent experiments. Scale bars, 10 μ m for low magnification, 2 μ m for high magnification. **b**, Cells were infected and stained as in **a** and analysed by fluorescence microscopy. Bacteria that co-localize with exofacial PS were scored for their accessibility to anti-*Listeria* antibodies present in the extracellular medium. Data show that the majority of PS⁺ bacteria are not accessible to anti-*Listeria* antibodies. Averages \pm s.d. for two independent experiments are shown ($n = 100$).



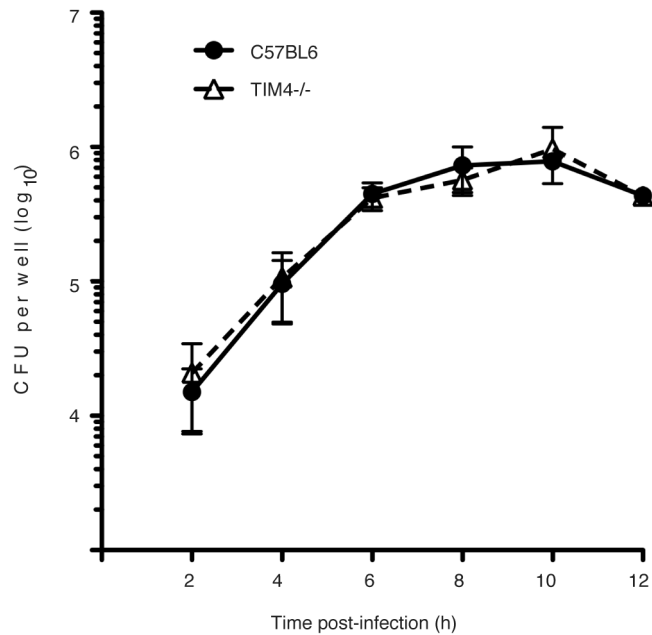
Extended Data Figure 6 | Formation of PS⁺ structures during *L. monocytogenes* infection of epithelial cells and macrophages.
a, Henle-407 human intestinal epithelial cells were infected with wild-type *L. monocytogenes* for 6 h and then incubated with a probe for exofacial PS (annexin A5–Alexa 488; green). Cells were then fixed and stained with

phalloidin to visualize F-actin (red) and bacteria (blue). Cells were analysed by fluorescence microscopy to identify PS⁺ structures and bacteria. Images are representative of three independent experiments. **b**, Mouse BMDMs from C57BL/6 mice were infected and stained as in **a**. Scale bars, 10 μm. Images are representative of three independent experiments.

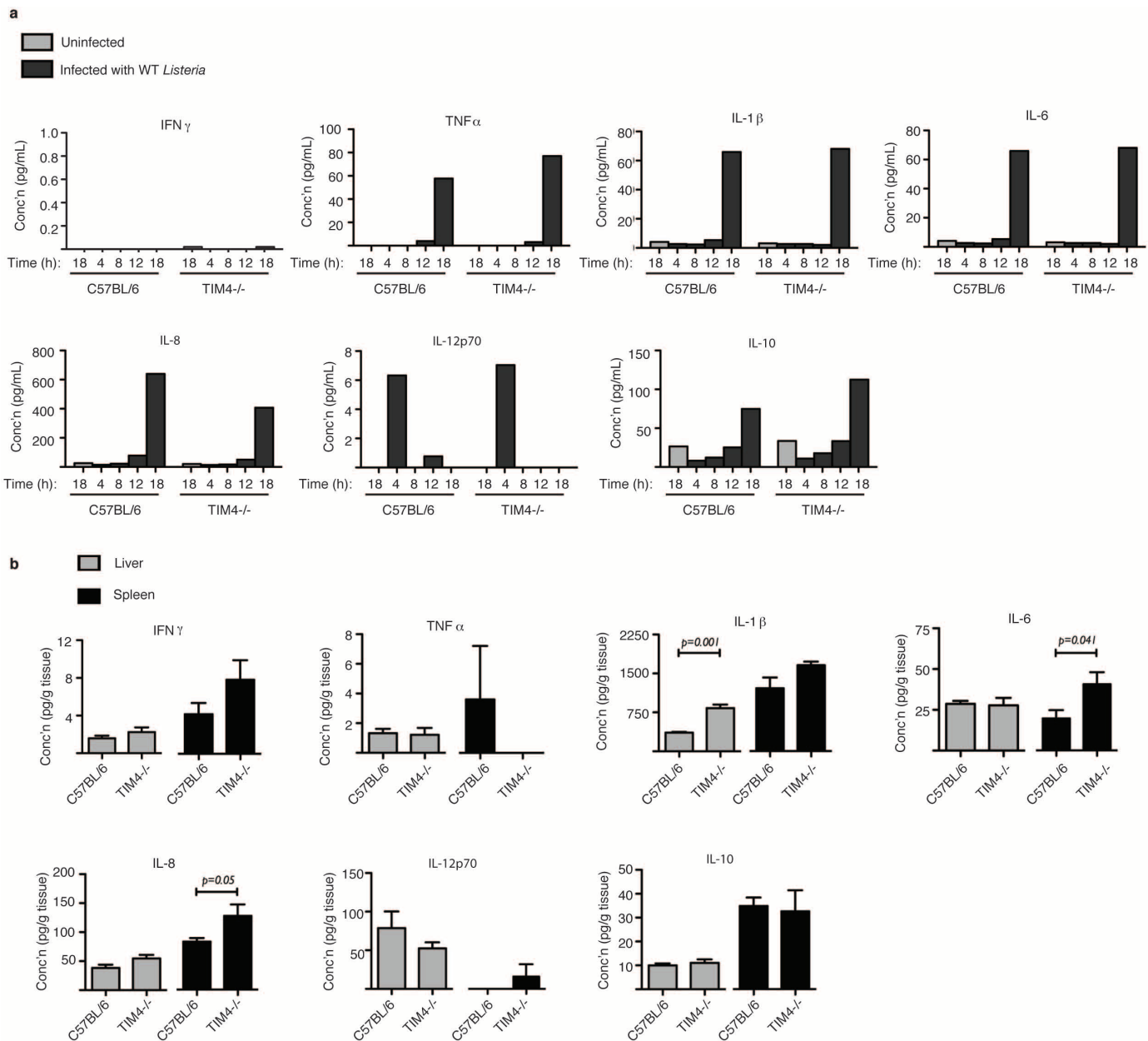


Extended Data Figure 7 | Release of PS⁺ structures containing *L. monocytogenes* from infected cells. **a**, HeLa cells were infected with wild-type *L. monocytogenes* for 6 h and then incubated with a probe for exofacial PS (annexin A5–Alexa 488; green). Cells were then fixed and stained with phalloidin to visualize F-actin (red) and bacteria (blue). Cells were analysed by fluorescence microscopy to identify PS⁺ structures and bacteria. Inset shows PS⁺ bacteria that are not cell associated. Images are representative of three independent experiments. **b**, HeLa cells were infected with $\Delta plcA\Delta plcB$

mutant bacteria for 6 h. The supernatant from the infected cultures was then removed and centrifuged onto poly-L-lysine-coated coverslips. Bacteria associated with coverslips were then stained with a probe for exofacial PS (annexin A5–Alexa 488; green). Cells were then fixed and stained for bacteria (blue). Coverslips were analysed by fluorescence microscopy to identify PS⁺ bacteria. Inset shows PS⁺ bacteria. Scale bars, 10 μ m for low-magnification images, 2 μ m for insets. Images are representative of three independent experiments.

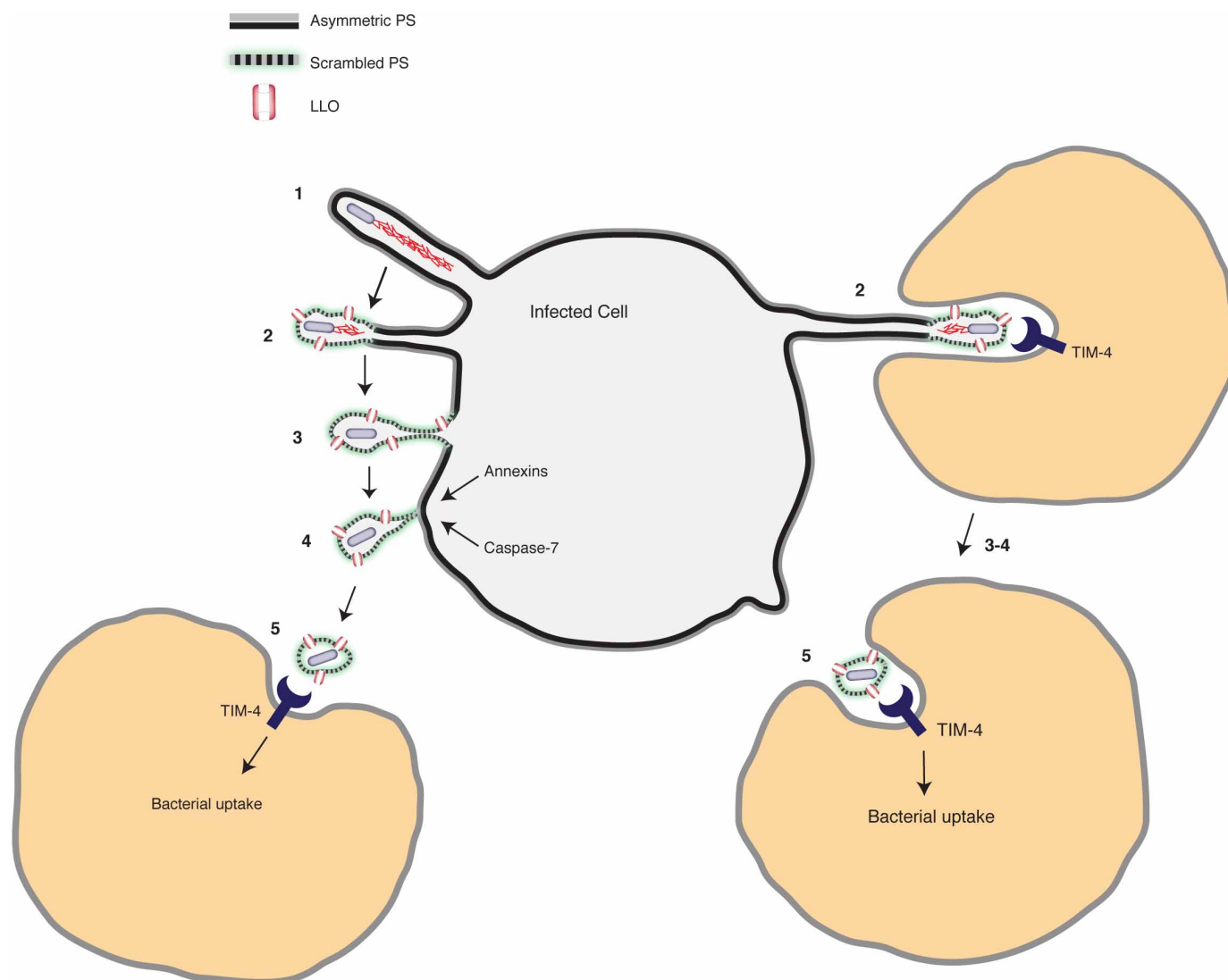


Extended Data Figure 8 | Growth of *L. monocytogenes* in the cytosol of *Timd4*^{-/-} macrophages is not impaired. Gentamicin protection assay to measure intracellular bacterial growth. BMDMs were harvested from C57BL/6 or *Timd4*^{-/-} mice and seeded at a density of 3×10^5 cells per well. Cells were then infected with wild-type *L. monocytogenes* in the presence of extracellular gentamicin. At the indicated times, cell lysates were plated and intracellular bacterial numbers (c.f.u.) were determined. Averages \pm s.d. for two independent experiments are shown.



Extended Data Figure 9 | Cytokine measurements. **a**, Measurement of cytokines after *in vitro* infection of BMDMs from C57BL/6 or *Timd4*^{-/-} mice with wild-type *L. monocytogenes* for the indicated time. Data from one of two independent experiments are shown. **b**, Measurement of basal cytokines in

tissues of C57BL/6 or *Timd4*^{-/-} mice without infection. Averages \pm s.d. for three independent experiments are shown. *P* values were calculated using one-way ANOVA.



Extended Data Figure 10 | *L. monocytogenes* exploits efferocytosis to promote cell-to-cell spread during infection. Model shows the steps that promote cell-to-cell spread by *L. monocytogenes*. **1.** Protrusion formation via actin-based motility. **2.** LLO-mediated damage to the plasma membrane leads to loss of membrane asymmetry and exofacial PS on protrusions. The exofacial exposure of PS promotes protrusion association with neighbouring cells (right). **3.** Loss of membrane asymmetry and PS exposure extends along length of protrusions. **4.** Calcium entry activates membrane repair pathways that promote scission of the protrusion. Bacteria are released from the cell in

PS⁺ vesicles. **5.** Macrophages mediate uptake of PS⁺ vesicles containing bacteria via the PS-binding receptor TIM-4. PS⁺ vesicles may be engulfed by neighbouring cells either near the infected cell surface (left side) or within enclosed spaces that form as a result of protrusion penetration into the neighbouring cell (right side). TIM-4 may also promote *L. monocytogenes* infection indirectly, through its ability to suppress basal levels of pro-inflammatory cytokines as part of its homeostatic function in the immune system.

NRROS negatively regulates reactive oxygen species during host defence and autoimmunity

Rajkumar Noubade^{1,†}, Kit Wong¹, Naruhisa Ota¹, Sascha Rutz¹, Celine Eidenschenk¹, Patricia A. Valdez^{1,†}, Jiabing Ding¹, Ivan Peng¹, Andrew Sebrell², Patrick Caplazi³, Jason DeVoss¹, Robert H. Soriano⁴, Tao Sai², Rongze Lu¹, Zora Modrusan⁴, Jason Hackney⁵ & Wenjun Ouyang¹

Reactive oxygen species (ROS) produced by phagocytes are essential for host defence against bacterial and fungal infections. Individuals with defective ROS production machinery develop chronic granulomatous disease^{1,2}. Conversely, excessive ROS can cause collateral tissue damage during inflammatory processes and therefore needs to be tightly regulated. Here we describe a protein, we termed negative regulator of ROS (NRROS), which limits ROS generation by phagocytes during inflammatory responses. NRROS expression in phagocytes can be repressed by inflammatory signals. NRROS-deficient phagocytes produce increased ROS upon inflammatory challenges, and mice lacking NRROS in their phagocytes show enhanced bactericidal activity against *Escherichia coli* and *Listeria monocytogenes*. Conversely, these mice develop severe experimental autoimmune encephalomyelitis owing to oxidative tissue damage in the central nervous system. Mechanistically, NRROS is localized to the endoplasmic reticulum, where it directly interacts with nascent NOX2 (also known as gp91^{phox} and encoded by *Cybb*) monomer, one of the membrane-bound subunits of the NADPH oxidase complex, and facilitates the degradation of NOX2 through the endoplasmic-reticulum-associated degradation pathway. Thus, NRROS provides a hitherto undefined mechanism for regulating ROS production—one that enables phagocytes to produce higher amounts of ROS, if required to control invading pathogens, while minimizing unwanted collateral tissue damage.

In response to microorganisms and inflammatory stimuli, professional phagocytes can generate ROS either within mitochondria or through a process named oxidative burst mediated by the NADPH oxidase 2 (NOX2) complex^{1–3}. Although many regulatory factors for ROS production in phagocytes have been described, our knowledge about its precise control is still limited^{1,4–6}. Although priming bone marrow-derived macrophages (BMDMs) with either interferon- γ (IFN- γ) or lipopolysaccharide (LPS) was sufficient to induce ROS upon subsequent zymosan stimulation, treatment with a combination of IFN- γ and LPS had a synergistic effect (Fig. 1a)^{7–9}. Although IFN- γ or LPS could enhance the expression of NOX2 (ref. 9), we speculated that there might be other mechanisms controlling this synergistic ROS production in phagocytes. Gene expression analysis by microarray under these conditions identified a previously uncharacterized gene, EMSMUSG 00000052384, which we named *Nrros* (negative regulator of ROS, previously known as *Lrrc33*) that was markedly downregulated upon priming with a combination of IFN- γ and LPS (Extended Data Fig. 1a). The predicted structure of NRROS contains a signal sequence, 21 putative leucine-rich repeat domains, a transmembrane domain, and a short cytoplasmic domain. NRROS was preferentially expressed in immune organs such as lymph nodes, thymus and spleen in mice (Extended Data Fig. 1b, c). Among leukocytes, myeloid cell subsets including macrophages, neutrophils and dendritic cells had the highest NRROS expression

(Fig. 1b and Extended Data Fig. 1d, e). Interestingly, priming with a combination of IFN- γ and LPS or tumour necrosis factor (TNF)- α alone markedly repressed *Nrros* messenger RNA and protein expression in wild-type BMDMs (Fig. 1c, d).

To reveal the biological functions of NRROS, we generated NRROS-specific antibody and NRROS-deficient mice (Extended Data Fig. 1f–j). At 6 weeks of age, all mice were viable and immune organs and leukocyte subsets were indistinguishable from those of wild-type mice (Extended Data Table 1 and data not shown). However, significantly augmented ROS production was observed from NRROS-deficient primary BMDMs upon zymosan stimulation after priming for 24 h with either IFN- γ (Fig. 1e) or LPS (Fig. 1f). These observations were confirmed in a variety of phagocytes, under several priming and activation

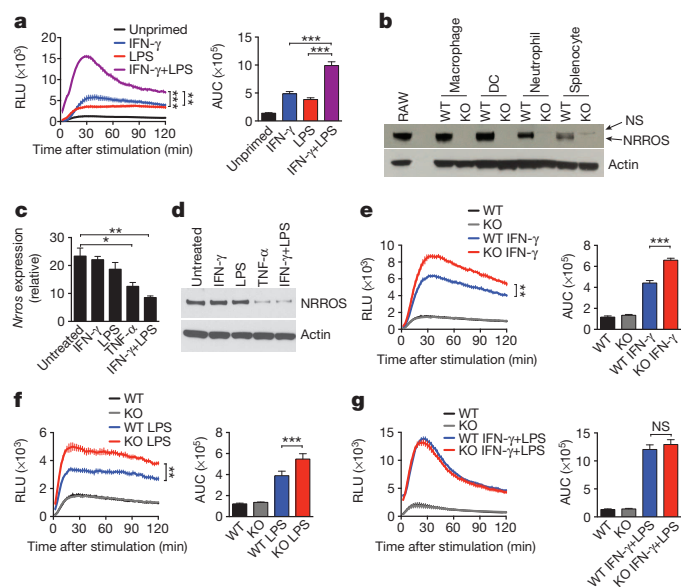


Figure 1 | NRROS is a negative regulator of ROS in phagocytes.

a, Zymosan-induced ROS production in BMDMs primed as indicated. **b**, NRROS expression analysis by western blot in immune cells. Actin was used as loading control. KO, NRROS-deficient knockout; NS, nonspecific band; WT, wild type. **c**, **d**, mRNA (**c**) and protein (**d**) of NRROS in BMDMs treated as indicated. **e–g**, Zymosan-induced ROS production in WT and KO BMDMs primed with IFN- γ (**e**), LPS (**f**) or both (**g**). **a**, **e–g**, Left, ROS kinetic plots of a representative experiment with three independent samples per group. Right, averaged area under curve (AUC) from five independent experiments. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Unpaired Student's *t*-test with Welch's corrections, (**a**, **e–g**, left), paired Student's *t*-test (**a**, **e–g**, right), unpaired Student's *t*-test (**c**). RLU, relative light units. Data in **b–d** represent at least three independent experiments.

¹Department of Immunology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ²Department of Antibody Engineering, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ³Department of Pathology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ⁴Department of Molecular Biology, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. ⁵Department of Bioinformatics, Genentech Inc., 1 DNA Way, South San Francisco, California 94080, USA. [†]Present addresses: Flexus Biosciences, 75 Shoreway Road, Suite D, San Carlos, California 94070, USA (R.N.); American Society for Biochemistry and Molecular Biology, 11200 Rockville Pike, Suite 302, Rockville, Maryland 20852, USA (P.A.V.).

conditions using multiple ROS detection methods (Extended Data Fig. 2). Interestingly, priming with a combination of IFN- γ and LPS, a condition that markedly repressed NRROS expression, attenuated the ROS-production difference between wild-type and NRROS-deficient cells (Fig. 1g), supporting the concept that inflammatory signals could control ROS production by regulating NRROS expression. Finally, NRROS exhibited a similar expression pattern and regulation in human cells (Extended Data Fig. 3a–c), and short interfering RNA (siRNA)-mediated knockdown of NRROS in human monocyte-derived macrophages also resulted in increased ROS production (Extended Data Fig. 3d–h). Taken together, these results suggested that NRROS functioned as a general negative regulator of ROS production in phagocytes.

Increased ROS production from NRROS-deficient BMDMs was also observed upon challenge with bacterial pathogens, including *E. coli*, *Salmonella enterica* serovar Typhimurium (hereafter *S. typhimurium*) and heat-killed *L. monocytogenes* (Fig. 2a–c and Extended Data Fig. 4a–c). More importantly, NRROS-deficient BMDMs and neutrophils were highly bactericidal *in vitro* (Fig. 2d and Extended Data Fig. 4d), which was not due to increased nitric oxide production or defective phagocytosis (Extended Data Fig. 4e, f). Conversely, enhanced bacterial killing by NRROS-deficient macrophages was eliminated upon pre-treatment

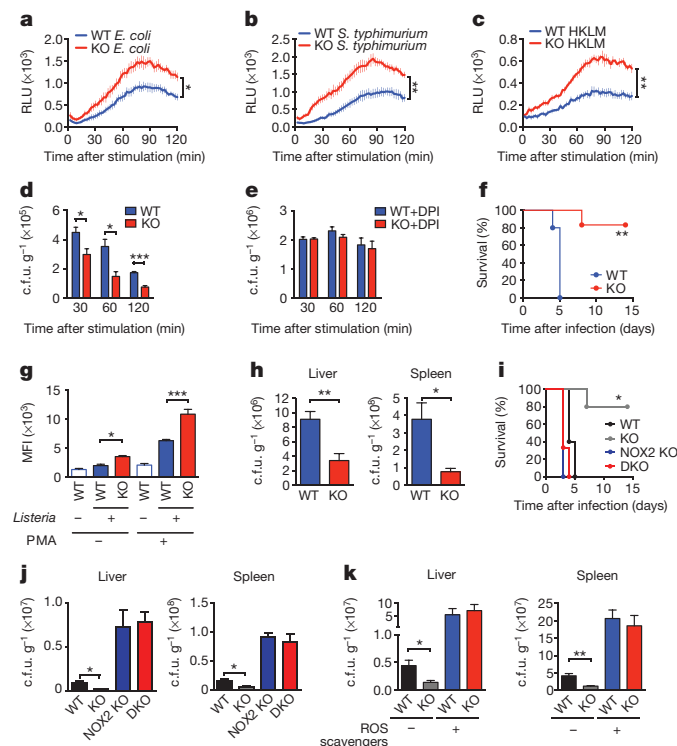


Figure 2 | Increased ROS and bactericidal ability in NRROS-deficient phagocytes. **a–c**, ROS production in WT and NRROS-deficient (KO) BMDMs stimulated *in vitro* with *E. coli* (**a**), *S. typhimurium* (**b**) or heat-killed *Listeria monocytogenes* (HKLM; **c**). **d, e**, *In vitro* *E. coli* killing by IFN- γ -primed WT and KO BMDMs without (**d**) or with (**e**) diphenyleneiodonium (DPI) pre-treatment. **f**, Survival curve of *L. monocytogenes*-infected WT ($n = 5$) and KO ($n = 6$) mice. **g**, *Ex vivo* ROS production by *L. monocytogenes*-infected WT and KO splenocytes ($n = 3$). **h**, *L. monocytogenes* burden in liver and spleen of WT ($n = 5$) and KO ($n = 6$) mice. **i**, Survival curve of *L. monocytogenes*-infected WT ($n = 5$), KO ($n = 6$), NOX2-deficient (NOX2 KO) ($n = 5$) and NRROS NOX2 double-deficient (DKO) ($n = 3$) mice. **j**, *L. monocytogenes* burden in liver and spleen of WT, KO, NOX2 and DKO mice ($n = 3$). **k**, *L. monocytogenes* burden in liver and spleen of WT and KO mice ($n = 5$) pre-treated with ROS scavenger cocktail. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Unpaired Student's *t*-test with (**a–c**) and without (**d, e, g, h, j, k**) Welch's corrections, log-rank (Mantel–Cox) test (**f, i**). c.f.u., colony-forming units. Data are representative of three (**a–f, h**) and two (**g, i–k**) independent experiments.

with diphenyleneiodonium, an inhibitor of ROS production (Fig. 2e). ROS is known to contribute to controlling *L. monocytogenes* infection, especially at higher bacterial burden^{10,11}. At a dose at which most wild-type mice succumbed, most NRROS-deficient mice survived the infection (Fig. 2f). Further, NRROS expression in splenic myeloid cells during infection, even though detectable, was highly repressed (Extended Data Fig. 4g). In accordance with our *in vitro* data, ROS production was significantly higher in spleens of infected NRROS-deficient mice (Fig. 2g). Importantly, survival of NRROS-deficient mice correlated with significantly reduced *L. monocytogenes* burden in livers and spleens of these mice (Fig. 2h). Similar enhanced bacterial killing was also observed upon inoculation of NRROS-deficient mice with live *E. coli* (data not shown). The protective effects observed in NRROS-deficient mice during *L. monocytogenes* infection were completely abrogated when the mice were crossed to NOX2-deficient mice or treated with ROS scavengers (Fig. 2i–k)¹², suggesting that increased ROS production in NRROS-deficient phagocytes contributed to the enhanced host defence against invading pathogens.

Despite the beneficial role in host defence, increased ROS production can also cause collateral tissue damage. As a negative regulator, NRROS might be required to control excessive ROS production during inflammation. Indeed, NRROS-deficient bone marrow chimaeric mice showed enhanced disease severity and high mortality in a myelin oligodendrocyte glycoprotein (MOG)_{35–55}-complete Freund's adjuvant (CFA)-induced experimental autoimmune encephalomyelitis (EAE) model (Fig. 3a, b). A similar high severity of EAE was observed in phagocyte-specific NRROS-deficient mice (*Nrros*^{fl/fl} LysM-Cre^{pos}) (Extended Data Fig. 5a–c), excluding a possible contribution of T cells, especially T_H17 cells, in this phenotype^{13,14}. Histological analysis revealed markedly increased lesion severity in *Nrros*^{fl/fl} LysM-Cre^{pos} mice compared to control mice (Extended Data Fig. 5d, e). Central nervous system (CNS)-infiltrating leukocytes—especially CD11b⁺ myeloid subsets—from NRROS-deficient bone marrow chimaeric mice showed significantly increased ROS production (Fig. 3c, d). Levels of malondialdehyde, one of the major toxic products of lipid peroxidation and indicator of oxidative damage¹⁵, were significantly higher in the CNS tissues from NRROS-deficient mice than in those from wild-type mice (Fig. 3e, f). Finally, treatment with ROS scavengers after EAE induction reduced disease severity and abrogated the difference between wild-type and NRROS-deficient bone marrow chimaeric mice (Fig. 3g) or *Nrros*^{fl/fl} LysM-Cre^{pos} mice (Extended Data Fig. 5f, g), supporting the premise that the increased disease severity in NRROS-deficient mice was caused by augmented tissue oxidation and damage. In summary, these data demonstrated an indispensable role for phagocyte-specific NRROS in limiting ROS-induced tissue damage under inflammatory conditions.

Next we investigated how NRROS regulates ROS production from myeloid cells. We did not detect any differences in mitochondrial ROS (mROS) generation despite an increased total ROS production in NRROS-deficient BMDMs treated with IFN- γ and zymosan (Fig. 4a) or with mROS-specific stimulants, such as antimycin A and rotenone (Extended Data Fig. 6a, b). Conversely, pharmacological inhibition of NOX2 but not of mROS production largely abrogated the differential ROS generation between wild-type and NRROS-deficient BMDMs (Extended Data Fig. 6c–g), implying that NRROS specifically regulates the function of the NOX2 complex. The NOX2 complex is composed of two membrane-bound subunits, NOX2 and p22^{phox} (encoded by *Cyba*) and four cytosolic subunits, p40^{phox}, p47^{phox}, p67^{phox} and RAC1 or RAC2^{1,2,8,16}. NOX2 and p22^{phox} form a heterodimer known as flavocytochrome *b*₅₅₈ in the endoplasmic reticulum (ER), and this heterodimerization is essential for the stability of these proteins^{17,18}. In NRROS-deficient BMDMs, the expression, phosphorylation and activation of the cytoplasmic subunits were comparable to those of wild-type BMDMs (Extended Data Fig. 7a–c). However, protein, but not mRNA, levels of NOX2 and p22^{phox} were increased in NRROS-deficient BMDMs (Fig. 4b and Extended Data Fig. 7d, e). Similarly, an increased expression of NOX2 and p22^{phox} was observed in various primary

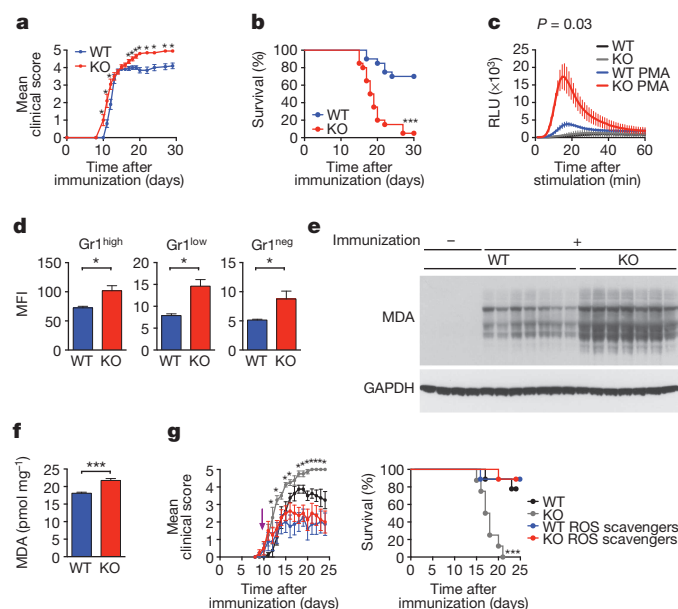


Figure 3 | Mice with NRROS-deficient haematopoietic cells develop severe EAE owing to oxidative damage in the CNS. **a, b**, Clinical course (**a**) and survival curve (**b**) of WT and NRROS-deficient (KO) bone marrow chimeric mice immunized with MOG_{35–55}-CFA ($n = 20$). **c**, ROS production in CNS cells from EAE-induced mice ($n = 4$ WT, 3 KO). **d**, ROS production in CD11b⁺ CNS cells from EAE-induced mice analysed by fluorescence-activated cell sorting ($n = 5$). MFI, mean fluorescence intensity. **e, f**, Malondialdehyde (MDA) adducts in the CNS tissues of EAE-induced mice analysed by western blot (**e**) and enzyme-linked immunosorbent assay (ELISA) (**f**) ($n = 7$). GAPDH as loading control. **g**, Clinical course (left) and survival curve (right) of EAE-induced mice ($n = 9$) treated with a ROS scavenger cocktail from day 10 (arrow). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Unpaired Student's *t*-test without (**a, d, f, g**) and with (**c**) Welch's corrections, log-rank (Mantel–Cox) test (**b, g**). Data are representative of two (**c–g**) and four (**a, b**) independent experiments.

phagocytes from NRROS-deficient mice after inflammatory challenges *in vivo* (Extended Data Fig. 7f–h), supporting a role for NRROS in regulating protein expression or stability of the flavocytochrome *b*₅₅₈ subunits in phagocytes.

To better understand how NRROS regulates the expression of NOX2 and p22^{phox} proteins, BMDMs were treated with cycloheximide to block *de novo* protein synthesis. NOX2 and p22^{phox} were markedly stabilized for extended durations in NRROS-deficient BMDMs compared to wild-type cells upon treatment (Fig. 4c), suggesting that NRROS participated in the degradation of these proteins. If not dimerized, NOX2 and p22^{phox} are presumed to be quickly degraded through a proteasome-dependent ER-associated degradation (ERAD) pathway^{17–19}. However, upon translocation out of the ER to phagosomal and plasma membranes, the NOX2–p22^{phox} complex might also get degraded by other mechanisms, such as the lysosome-associated degradation pathway. Interestingly, the proteasome inhibitor MG132, but not the lysosome blocker chloroquine, normalized the differences in NOX2 or p22^{phox} protein expression between wild-type and NRROS-deficient BMDMs (Fig. 4d, e and Extended Data Fig. 8a–e). Subcellular fractionation revealed that the accumulation of NOX2 and p22^{phox} in NRROS-deficient cells started in the ER (Fig. 4f). In addition, most NOX2 detected in BMDMs was the 58-kDa ER-bound form (Fig. 4f)²⁰, which was sensitive to both Peptide-*N*-glycosidase F (PNGase F) and Endoglycosidase H (Endo H) enzymes (Fig. 4g)²¹. Finally, when newly synthesized NOX2 was metabolically labelled with [³⁵S]methionine, higher NOX2 levels were already evident as early as the end of 1-h pulse period (time 0) and a slower decay of NOX2 persisted for up to 4 h during the chase period in NRROS-deficient BMDMs (Fig. 4h). This increased NOX2 was Endo H sensitive (Extended Data Fig. 8f), demonstrating its ER origin. In

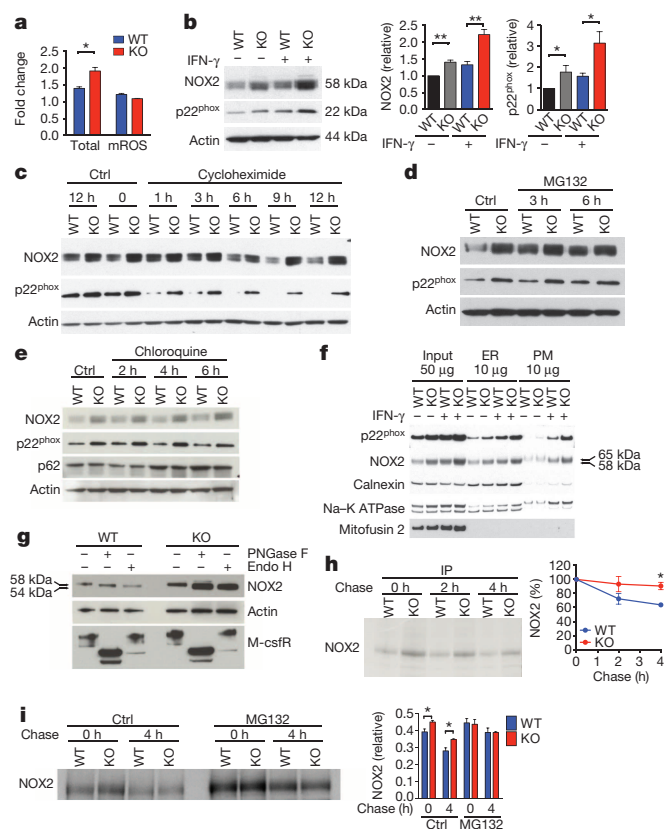


Figure 4 | NRROS regulates the activity of NOX2 oxidase complex in BMDMs. **a**, Zymosan-induced total ROS and mitochondrial ROS (mROS) in IFN- γ -primed WT and NRROS-deficient (KO) BMDMs. Fold change relative to unstimulated control. **b**, Left, NOX2 and p22^{phox} in unprimed (–) and IFN- γ -primed (+) BMDMs. Actin was used as loading control. Right, densitometry analyses relative to actin ($n = 4$). **c–e**, NOX2 and p22^{phox} in IFN- γ -primed BMDMs treated with cycloheximide (**c**), MG132 (**d**) or chloroquine (**e**). **f**, Subcellular fractionation and blotting of unprimed (–) and IFN- γ -primed (+) BMDMs. PM, plasma membrane. **g**, NOX2 in PNGase- or Endo H-treated IFN- γ -primed BMDM lysates. M-csfR was used as control for Endo H resistance. Actin was used as loading control. **h, i**, Left, autoradiography of immunoprecipitated (IP) NOX2 from [³⁵S]methionine-labelled IFN- γ -primed BMDMs without (**h**) or with (**i**) MG132 treatment. Right, densitometry analyses relative to control in the flow-through with (**h**) or without (**i**) normalization to 0 h samples ($n = 3$). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$ (unpaired Student's *t*-test). Data are representative of two (**a, g**) and at least three (**b–f, h–i**) independent experiments.

addition, the difference in protein degradation was abrogated when cells were treated with MG132 (Fig. 4i), suggesting that NRROS facilitated the degradation of nascent NOX2 through the ERAD pathway in phagocytes.

Consistent with these results, we demonstrated that NRROS mostly resided in the ER by subcellular fractionation, flow cytometric analysis and immunofluorescent staining (Fig. 5a and Extended Data Fig. 8g, h). Recent studies have suggested that the heat shock protein (HSP)70–HSP90–CHIP axis is involved in the degradation of NOX2 through ERAD^{19,22}. However, expression of HSP70, HSP90 and CHIP were comparable between wild-type and NRROS-deficient macrophages (Extended Data Fig. 9a). Even though pre-treatment with the HSP90 inhibitor radicicol markedly inhibited NOX2 expression in both wild-type and NRROS-deficient cells—almost to a level where the difference in ROS production between wild-type and NRROS-deficient BMDMs was no longer observed (Extended Data Fig. 9b, c)—the higher NOX2 levels in NRROS-deficient BMDMs were maintained (Extended Data Fig. 9d). These data suggested that NRROS most likely controls NOX2 degradation in the ER by a different mechanism, although we could not exclude a possible cooperation with the HSP90–HSP70–CHIP axis.

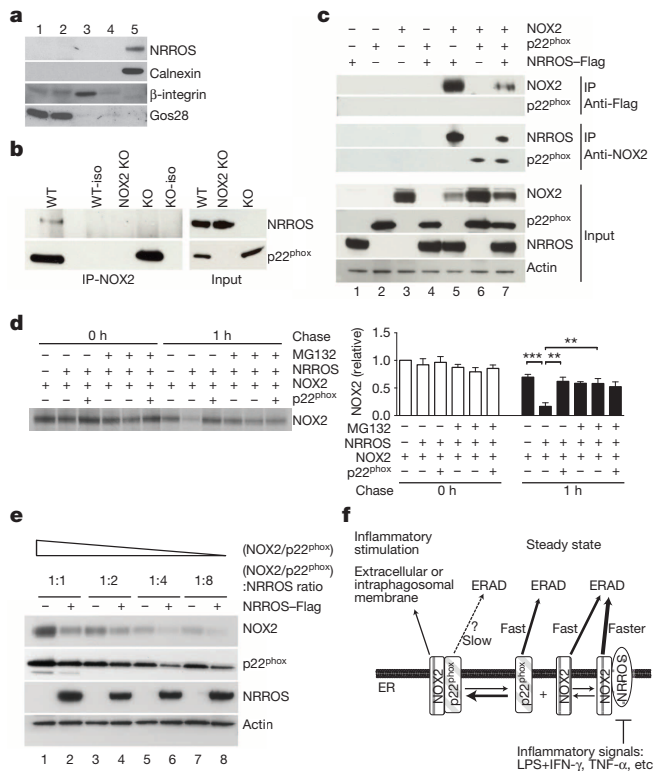


Figure 5 | NRROS promotes NOX2 degradation in the ER. **a**, Subcellular fractionation of RAW cells and western blot analysis. Numbers represent fractions. **b**, Anti-NOX2-immunoprecipitated NRROS and p22^{phox} from IFN- γ -primed BMDMs. iso, isotype control. **c**, NRROS-Flag, NOX2 and p22^{phox} immunoprecipitated with Flag or NOX2 antibody from CHO cells transfected with the indicated plasmids. Bottom panels, input samples. Actin was used as loading control. Numbers below indicate lanes. **d**, Autoradiography of NOX2 from [³⁵S]methionine-labelled, MG132-treated 293T cells transfected with the indicated plasmids. Right graph, densitometry relative to time 0 of cells transfected with NOX2 alone. **e**, NOX2 and p22^{phox} in CHO cells transfected with the indicated ratio of NOX2/p22^{phox} (1:1)-expressing plasmids to NRROS-Flag-expressing plasmid. Actin was used as loading control. Numbers below indicate lanes. **f**, Model depicting NOX2 complex regulation by NRROS. ** $P < 0.01$, *** $P < 0.001$ (unpaired Student's t -test). Data are representative of at least three independent experiments.

Next, given their colocalization in the ER, we examined whether NRROS interacts with the NOX2-p22^{phox} complex. Interestingly, co-immunoprecipitation of NRROS by NOX2 was detected in wild-type BMDMs (Fig. 5b). As neither the NOX2 nor the p22^{phox} monomer is stable in primary phagocytes, we further studied their interaction with NRROS in CHO and 293T cells²³. Surprisingly, upon ectopic over-expression of Flag-tagged NRROS, together with NOX2 and/or p22^{phox} in CHO cells, we observed a specific interaction between NRROS and NOX2 only, and not between NRROS and the p22^{phox} monomer or the NOX2-p22^{phox} heterodimer (Fig. 5c, lanes 5–7, top panel). On the other hand, NOX2 could form a heterodimer with either NRROS or p22^{phox} (Fig. 5c, lanes 5–7, middle panel). Importantly, p22^{phox} interfered with the interaction between NRROS and NOX2 (Fig. 5c, lanes 5 and 7, top and middle panels). Co-expression of NOX2 with NRROS led to a substantial downregulation of NOX2 protein (Fig. 5c, lanes 3 and 5, bottom panel), but not the transcript (data not shown). The expression of p22^{phox} protein remained unchanged or was only mildly reduced upon NRROS co-expression (Fig. 5c, lanes 2 and 4, bottom panel). Consistent with the co-immunoprecipitation data, the formation of a NOX2-p22^{phox} heterodimer prevented the NRROS-mediated downregulation of NOX2 (Fig. 5c, lanes 5 and 7, bottom panels). Similar results were also obtained in 293T cells (data not shown). Pulse chase experiments in 293T cells further suggested that NRROS facilitated the degradation of NOX2

through the ERAD pathway, which could be prevented either by MG132 treatment or co-expression of p22^{phox} (Fig. 5d). Finally, a relatively stable NOX2-p22^{phox} complex could be shifted to the NRROS-mediated degradation pathway with increased molar ratio of NRROS to NOX2-p22^{phox}, where in addition to the decline in NOX2 levels, a reduction in p22^{phox} levels was observed as well (Fig. 5e, lanes 5 to 8).

Altogether, our data support a model in which NRROS directly interacts with nascent NOX2 monomers and facilitates their degradation through the ERAD pathway (Fig. 5f)²⁴. p22^{phox} competes with NRROS to form a much more stable NOX2-p22^{phox} complex. Certain inflammatory signals can regulate the expression of NRROS, and thereby modulate the expression of the NOX2 complex and ROS production (Fig. 5f). The delicate regulation of NOX2 by NRROS ensures that there is sufficient ROS production by phagocytes during host defence while at the same time limiting collateral tissue damage. Further understanding of this pathway may provide novel therapeutic approaches to target ROS production in various diseases.

METHODS SUMMARY

All animal experiments were approved by the Genentech Institutional Animal Care and Use Committee (IACUC). Mice were infected with *L. monocytogenes* and monitored for mortality or euthanized for bacterial burden analyses. EAE was induced by immunizing mice with MOG_{35–55}-CFA/pertussis toxin and monitored for clinical score and mortality. Oxidative damage in the CNS of EAE-induced mice was assessed by malondialdehyde measurement with ELISA or western blot (OxiSelect MDA kit, Cell Biolabs). *In vivo* ROS neutralization was achieved by injecting mice with a cocktail of scavengers (15 mg kg⁻¹ EUK134 (Cayman Chemical), 20 mg kg⁻¹ FeTPPS (Calbiochem) and 50 mg kg⁻¹ PBN (Sigma-Aldrich)). *In vitro* *E. coli* killing was assessed by gentamicin protection assay. *In vitro*, ROS was measured by incubating primed phagocytes with an inducing agent and either luminol for continuous detection by luminometer or CM-H₂DCFDA (total ROS) or MitoSOX (mROS) (Invitrogen) for FACS analyses. *Ex vivo* ROS was measured by incubating cells with CM-H₂DCFDA and analysed by FACS. Whole cell/tissue lysates prepared in RIPA or Triton X-100 buffer were used for immunoprecipitation and western blot studies. Metabolic labelling of cells was performed using [³⁵S]methionine. Subcellular fractionation was carried out by sucrose gradient ultracentrifugation. Total RNA was used in microarray or Taqman expression analyses.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 14 December 2012; accepted 14 February 2014.

Published online 13 April 2014.

- Bedard, K. & Krause, K. H. The NOX family of ROS-generating NADPH oxidases: physiology and pathophysiology. *Physiol. Rev.* **87**, 245–313 (2007).
- Lambeth, J. D. NOX enzymes and the biology of reactive oxygen. *Nature Rev. Immunol.* **4**, 181–189 (2004).
- West, A. P. et al. TLR signalling augments macrophage bactericidal activity through mitochondrial ROS. *Nature* **472**, 476–480 (2011).
- Honda, F. et al. The kinase Btk negatively regulates the production of reactive oxygen species and stimulation-induced apoptosis in human neutrophils. *Nature Immunol.* **13**, 369–378 (2012).
- Wang, G. et al. Cutting edge: Slamf8 is a negative regulator of Nox2 activity in macrophages. *J. Immunol.* **188**, 5829–5832 (2012).
- Di, A. et al. The redox-sensitive cation channel TRPM2 modulates phagocyte ROS production and inflammation. *Nature Immunol.* **13**, 29–34 (2012).
- McPhail, L. C., Clayton, C. C. & Snyderman, R. The NADPH oxidase of human polymorphonuclear leukocytes. Evidence for regulation by multiple signals. *J. Biol. Chem.* **259**, 5768–5775 (1984).
- El-Benna, J., Dang, P. M. & Gougerot-Pocidalo, M. A. Priming of the neutrophil NADPH oxidase activation: role of p47phox phosphorylation and NOX2 mobilization to the plasma membrane. *Semin. Immunopathol.* **30**, 279–289 (2008).
- Anrather, J., Racchumi, G. & Iadecola, C. NF- κ B regulates phagocytic NADPH oxidase by inducing the expression of gp91^{phox}. *J. Biol. Chem.* **281**, 5657–5667 (2006).
- Dinauer, M. C., Deck, M. B. & Unanue, E. R. Mice lacking reduced nicotinamide adenine dinucleotide phosphate oxidase activity show increased susceptibility to early infection with *Listeria monocytogenes*. *J. Immunol.* **158**, 5581–5583 (1997).
- Shiloh, M. U. et al. Phenotype of mice and macrophages deficient in both phagocyte oxidase and inducible nitric oxide synthase. *Immunity* **10**, 29–38 (1999).
- Nikic, I. et al. A reversible form of axon damage in experimental autoimmune encephalomyelitis and multiple sclerosis. *Nature Med.* **17**, 495–499 (2011).

13. Tse, H. M. *et al.* NADPH oxidase deficiency regulates Th lineage commitment and modulates autoimmunity. *J. Immunol.* **185**, 5247–5258 (2010).
14. Ouyang, W., Kolls, J. K. & Zheng, Y. The biological functions of T helper 17 cell effector cytokines in inflammation. *Immunity* **28**, 454–467 (2008).
15. Wällberg, M., Bergquist, J., Achour, A., Breij, E. & Harris, R. A. Malondialdehyde modification of myelin oligodendrocyte glycoprotein leads to increased immunogenicity and encephalitogenicity. *Eur. J. Immunol.* **37**, 1986–1995 (2007).
16. Sheppard, F. R. *et al.* Structural organization of the neutrophil NADPH oxidase: phosphorylation and translocation during priming and activation. *J. Leukoc. Biol.* **78**, 1025–1042 (2005).
17. DeLeo, F. R. *et al.* Processing and maturation of flavocytochrome b558 include incorporation of heme as a prerequisite for heterodimer assembly. *J. Biol. Chem.* **275**, 13986–13993 (2000).
18. Parkos, C. A., Dinauer, M. C., Jesaitis, A. J., Orkin, S. H. & Curnutte, J. T. Absence of both the 91kD and 22kD subunits of human neutrophil cytochrome b in two genetic forms of chronic granulomatous disease. *Blood* **73**, 1416–1420 (1989).
19. Chen, F. *et al.* Hsp90 regulates NADPH oxidase activity and is necessary for superoxide but not hydrogen peroxide production. *Antioxid. Redox Signal.* **14**, 2107–2119 (2011).
20. Pollock, J. D. *et al.* Mouse model of X-linked chronic granulomatous disease, an inherited defect in phagocyte superoxide production. *Nature Genet.* **9**, 202–209 (1995).
21. Lewis, V. *et al.* Glycoproteins of the lysosomal membrane. *J. Cell Biol.* **100**, 1839–1847 (1985).
22. Chen, F. *et al.* Opposing actions of heat shock protein 90 and 70 regulate nicotinamide adenine dinucleotide phosphate oxidase stability and reactive oxygen species production. *Arterioscler. Thromb. Vasc. Biol.* **32**, 2989–2999 (2012).
23. Casbon, A. J., Allen, L. A., Dunn, K. W. & Dinauer, M. C. Macrophage NADPH oxidase flavocytochrome B localizes to the plasma membrane and Rab11-positive recycling endosomes. *J. Immunol.* **182**, 2325–2339 (2009).
24. Brodsky, J. L. Cleaning up: ER-associated degradation to the rescue. *Cell* **151**, 1163–1167 (2012).

Supplementary Information is available in the online version of the paper.

Acknowledgements We thank D. Seshasayee and H. Van Gorp for providing L929 condition media; J. Tao for helping with the screening of NRROS antibodies; C. Kuijl, J. R. Junutula (Genentech) and D. Monack (Stanford University) for providing *S. typhimurium*; W. Lee for providing *L. monocytogenes* and K. Huynh for providing dsRed *E.coli*; W. Forrest (Genentech) for help with statistical analysis; M. Roose-Girma, M. Schlatter and J. Aurellano (Genentech) for help with design and generation of *Nrros* conditional knockout construct and ES cells.

Author Contributions W.O. and R.N. devised the project and wrote the manuscript. R.N. designed and performed most of the experiments. K.W. contributed to Fig. 2h and Extended Data Figs 1e, 2a, b. N.O. contributed to Fig. 2f. S.R. contributed to Fig. 3a–g. C.E. contributed to Extended Data Fig. 5c–g. P.A.V. contributed to Extended Data Table 1, R.L. contributed to Fig. 5c, d. K.W., N.O., S.R. and C.E. helped to edit the manuscript. J.D. made the constructs used in Fig. 5 and Extended Data Fig. 8g, h. I.P. and J.D.V. contributed to EAE experiments in Fig. 3a, b, g and Extended Data Fig. 5c, f. A.S. and T.S. assisted in generating NRROS-specific antibodies (Fig. 1b and Extended Data Fig. 1c, e, h–j). P.C. carried out histopathology analyses and contributed to Extended Data Fig. 5d, e, g. R.H.S. and Z.M. assisted in microarray analysis and J.H. performed bioinformatics analysis (Extended Data Fig. 1a).

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare competing financial interests: details are available in the online version of the paper. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to W.O. (ouyang.wenjun@gene.com).

METHODS

Mice. NRROS-deficient mice were generated in collaboration with Lexicon Pharmaceuticals²⁵ and backcrossed to C57BL/6J more than 14 generations. NOX2-deficient mice were procured from Jackson Laboratories and bred with NRROS-deficient mice to generate NRROS NOX2 double-deficient mice. NRROS conditional knockout (*Nrros*^{fl/fl}) mice were generated in-house by targeting exon 4 of *Nrros* that encodes most of the protein. These floxed mice were bred with LysM-Cre mice (kind gift from University of California, San Diego) to specifically delete NRROS expression in M lysozyme-expressing cells, namely macrophages, monocytes and neutrophils. Mice were routinely genotyped by PCR using the following primers: 5'-TACTCACTGGCCTTGCCTC-3', 5'-AGAGTTCCATCCTCAGTCCC-3' and 5'-GTTACATCTGGTGCCAGAC-3'. C57BL/6J and B6.SJL-*Ptprca*^a*Pepec*^b/BoyJ mice were procured from Jackson Laboratories. *Rag2*-deficient mice (B6.129S6-*Rag2*^{tm1Fwa} N12) were obtained from Taconic.

For the generation of bone marrow chimaeras, B6.SJL-*Ptprca*^a*Pepec*^b/BoyJ recipient mice were irradiated with a total of 1,050 rads (in two doses of 525 rads separated by 4-h interval) using a ¹³⁷Cs source. 10–15 × 10⁶ total bone marrow cells isolated from femur and tibias of the wild-type and NRROS-deficient donor mice were injected into the tail vein. Reconstituted mice were given water containing 0.11 mg ml⁻¹ polymyxin B and 1.1 mg ml⁻¹ neomycin for 2 weeks and then switched to regular water. Mice were allowed to reconstitute for at least 12 weeks (more than 90% reconstitution as assessed by CD45.1 (A20, catalogue no. 553776) and CD45.2 antibody (104, catalogue no. 558702) (BD Biosciences) staining and FACS (LSR II, BD Biosciences)) and used for *in vivo* experiments. All animal experiments were approved by the Genentech Institutional Animal Care and Use Committee.

Isolation of cells. For BMDMs, femur and tibia were collected from 3–5 mice of each genotype and bone marrow cells were flushed with complete DMEM containing 50 U ml⁻¹ penicillin, 50 mg ml⁻¹ streptomycin, 2 mM L-glutamine, sodium pyruvate and 10% FBS. Erythrocytes were removed by ACK red blood cell lysis buffer treatment and the cell suspension was filtered through a 70-µm cell strainer to remove any cell clumps. The single-cell suspensions were then cultured at 37 °C for 1 h, and non-adherent cells were collected and re-plated in complete DMEM with 20% L929 conditioned media. To fully differentiate BMDMs, the cells were cultured for an additional 6 days with media change every 3 days. All the cells, when analysed by FACS, were CD11b⁺F4/80⁺ (BD Biosciences). In some experiments, recombinant mouse M-CSF (R&D Systems) was used at a concentration of 50 ng ml⁻¹ instead of L929 medium. BMDMs were primed with 10 ng ml⁻¹ of IFN-γ (R&D Systems) for 18–24 h, unless otherwise stated.

For peritoneal macrophages, mice were injected with 1 ml of 4% thioglycollate, and 4 days later peritoneal cells were collected by lavage using 8 ml PBS. Cells were allowed to adhere for at least 2 h and adherent macrophages were collected, counted and plated in complete DMEM for further analysis.

For neutrophil isolation, single-cell suspensions of the total bone marrow cells from 3–4 mice were prepared and subjected to Percoll gradient separation (78–69%–52%) by centrifugation at 1,500g for 40 min. Neutrophils were collected from 78–69% interphase, washed twice, re-suspended in HBSS and used for further analysis. Cell purity was more than 90% as analysed by FACS staining with CD11b (M1/70, catalogue no. 48-0112; eBioscience) and Gr-1 antibodies (RB6-8C5, catalogue no. 553126; BD Biosciences).

For studies with human cells, blood received at Genentech health centre from healthy donors was used to isolate peripheral blood mononuclear cells (PBMCs) by Ficoll gradient (GE Healthcare). Monocytes were further isolated from the PBMCs using human monocyte isolation kit II (Miltenyi Biotec) according to the manufacturer's instructions. Monocyte-derived macrophages were generated by culturing the purified monocytes for 7 days in complete RPMI 1640 containing 10% FBS and 100 ng ml⁻¹ human M-CSF (R&D Systems). Fresh medium containing 100 ng ml⁻¹ of M-CSF was added to cells on day 4 and cells were used for further analysis on day 7. PBMCs were also used to isolate B cells and T cells using CD19 microbeads and Pan T Cell Isolation Kit II (Miltenyi Biotec), respectively, according to the manufacturer's instructions.

Isolation of CNS-infiltrating mononuclear cells. Animals were perfused with 30 ml of PBS, and brain and spinal cords were collected. A single-cell suspension was obtained by grinding the tissues using 1 ml syringe plunger and passed through a 70-µm strainer. Mononuclear cells were obtained by Percoll gradient (37/70%, GE Healthcare) centrifugation at 390g for 30 min at room temperature (20–25 °C) collected from the interphase, washed and used for further analysis.

Plasmids. Sequence coding murine NRROS protein was amplified by PCR using mouse splenic complementary DNA as template, and inserted into BglIII(blunt)/EcoRI site of Nflag-pRK vector which has signal peptide sequence and Flag-tag sequence. Then, cDNA encoding signal peptide-Flag-tag-NRROS was transferred onto retroviral vector pMSCV-IRES-GFP to generate Flag-NRROS-pMSCV. This amino-terminally Flag-tagged NRROS (Flag-NRROS) was used to generate RAW stable cell lines. A plasmid expressing carboxy-terminally Flag-tagged NRROS

(NRROS-Flag) was generated by cloning PCR-amplified coding region of NRROS including its signal sequence into pRK5 vector using AscI and XbaI sites to generate NRROS-Flag-pRK. Mouse *Cyba* and *Cybb* genes encoding p22^{phox} and NOX2, respectively, were amplified by PCR using mouse splenic cDNA and cloned into pRK5 vector using EcoRI and XbaI sites to generate p22^{phox}-pRK and NOX2-pRK, respectively.

Generation of Flag-NRROS-RAW cells by retroviral transduction. Phoenix E cells were transfected with either Flag-NRROS-pMSCV or pMSCV-IRES-GFP control plasmid by the calcium phosphate method. After 48 h, the retroviral supernatants were collected, filtered through 0.45-µm filters, supplemented with polybrene (10 µg ml⁻¹) and added to RAW cells that were plated in a six-well plate the previous night. Plates were centrifuged at 1,200g at 32 °C for 120 min. The viral supernatant was replaced by complete DMEM after an additional 2 h. Transduced RAW cells were sorted (FACSARIA, BD Biosciences) based on GFP expression.

Transfection of cells. CHO cells were transfected with the indicated plasmid using Lipofectamine 2000 (Life Technologies) according to manufacturer's instructions. In brief, 10 µg total DNA was diluted in 500 µl of opti-MEM and 25 µl of Lipofectamine 2000 was diluted in another 500 µl of opti-MEM. The two were combined, incubated for 20 min and added drop-wise to cells grown on 100-mm plates at 70–80% confluency. After 24 h, cells were collected by trypsinization and lysed in RIPA buffer (50 mM Tris, pH 7.4, 150 mM NaCl, 2 mM EDTA, 1% NP-40, 1% SDS) and subjected to western blot analysis or were used for immunoprecipitation with the indicated antibody. For the titration experiments, 2.5 µg each of NOX2-pRK and p22^{phox}-pRK plasmids were mixed and then serially diluted two-fold while maintaining their 1:1 ratio. The amount of NRROS-Flag-pRK plasmid was held constant at 2.5 µg and added to the above tubes so that the ratio of NOX2-pRK:p22^{phox}-pRK:NRROS-Flag-pRK was 1:1:1, 1:1:2, 1:1:4 and 1:1:8. Total amount of DNA was adjusted to 10 µg with empty pRK plasmid. After 24 h cells were collected, lysed and subjected to western blot analyses. 293T cells grown in 100-mm plates were transfected using Fugene 6 reagent (Promega) at 3:1 ratio (DNA:Fugene), according to the manufacturer's instructions and processed as described for CHO cells. Both CHO and 293T cells were obtained from an internal cell-banking program (gCell) at Genentech.

siRNA knockdown. Chemically modified siRNAs²⁶ were synthesized in house. siRNA specific for *Nrros* was composed of 5'-mAmGAGCUAGAUUUGCAGAGdGdAdAmAmCmU-3' (sense) and 5'-dTUUCUCUGCAAAUCUAGCmCmU-3' (antisense), and the control scrambled siRNA was 5'-mGmGAGCGACCAU CUUCdTcdAmAmUmU-3' (sense) and 5'-dTUGAGAAGAUGGUGCGCmCmC-3' (antisense). RAW cells were transfected with ~600 pmol siRNA using Cell Line Nucleofection Kit V (Lonza). 24 h later, cells were re-plated at 5 × 10⁴ cells per well of a 96-well clear-bottom white plate (Corning), primed with 10 ng ml⁻¹ IFN-γ (R&D Systems) for another 24 h and used for further analysis.

ROS measurement. For luminol-based ROS measurement, 1 × 10⁵ macrophages were plated in clear-bottomed 96-well white plates (Corning) in triplicate, and primed for 24 h with 10 ng ml⁻¹ of IFN-γ, 100 ng ml⁻¹ LPS (Invivogen) or 100 ng ml⁻¹ TNF-α (R&D Systems), either alone or in combination. In most of the experiments, cells were primed with 10 ng ml⁻¹ IFN-γ unless otherwise indicated. Oxidative burst was induced in luminol medium (PBS containing 5 mM glucose, 1 mM MgCl₂, 0.5 mM CaCl₂ and 100 µM luminol (Sigma-Aldrich)) by stimulating macrophages with 100 µg ml⁻¹ zymosan A (Invivogen), 100 ng ml⁻¹ PMA (Sigma-Aldrich), heat-killed *Listeria monocytogenes* (Invivogen) at multiplicity of infection (m.o.i.) 50, *E. coli* strain 26 (ATCC) at m.o.i. 50, or *S. typhimurium* SL1344 (a kind gift from D. Monack, Stanford University) at m.o.i. 10. Luminescence was measured by Glomax luminometer (Promega). Neutrophils and CNS mononuclear cells were re-suspended in the luminol medium (1 × 10⁵ cells per well) and stimulated with 100 ng ml⁻¹ PMA or 100 µg ml⁻¹ zymosan. For experiments with inhibitors, the following chemicals were added to the cell cultures during the last 1 h of the priming: 5 µM diphenyleneiodonium (Calbiochem), 20 µM antimycin A (Sigma-Aldrich) or 3 mM apocynin (Calbiochem) unless otherwise indicated. For treatment with apocynin, the compound continued to be present during the stimulation, whereas for other treatments, cells were washed with PBS before ROS induction.

For FACS-based measurements, macrophages plated in 6-well plates were treated with the indicated stimulant for 30 min at 37 °C in the presence or absence of 5 µM CM-H₂DCFDA (Invitrogen) to measure total intracellular ROS or 5 µM MitoSOX (Invitrogen) to measure mitochondrial ROS. Cells were washed with PBS, collected by trypsinization and analysed by FACS. For ROS measurement induced by rotenone (Calbiochem) or antimycin A, cells were treated with or without 10 µM of the compounds for 2 h, washed with PBS, collected by trypsinization and analysed by FACS. Splenocytes and CNS mononuclear cells were stimulated with 100 ng ml⁻¹ PMA in the presence or absence of 5 µM CM-H₂DCFDA for 30 min, washed with PBS, collected and stained for surface markers CD11b (eBioscience), Gr-1, B220 and CD3 (BD Biosciences) for 30 min on ice before analysing by FACS.

Measurement of nitric oxide. BMDMs were treated with 10 ng ml^{-1} IFN- γ , 10 ng ml^{-1} TNF- α , 100 ng ml^{-1} LPS or their combination for 24 h, and nitric oxide in the culture supernatant was measured using Griess assay (Promega) with Spectra Max 340 (Molecular Devices) spectrophotometer according to the manufacturer's recommendations.

RNA isolation and real-time RT-PCR. RNA was isolated using RNeasy kit (Qiagen). Total RNA of 20 different human tissues was obtained from Clontech. Mouse major tissues total RNA was procured from Zygen. RNA samples were analysed by real-time RT-PCR with TaqMan One-Step RT-PCR Master Mix reagents (Applied Biosystems) and the appropriate primers and probes. Results were normalized to those of the housekeeping gene *Rpl19* (encoding ribosomal protein L19) and are reported as $2^{\Delta\text{CT}}$. The primer and probe sets used were: *Rpl19*: forward 5'-GCATCTCATGGAGCACAT-3', reverse 5'-CTGGTCAGCCAGGAGCTT-3', probe 5'-CTTGCGGGCCTTGTCTGCCTT-3'; *Nrros*: forward 5'-ACTGCAGC TTCCAAGGA-3', reverse 5'-TGGGTACCGAAGCAAGGT-3', probe 5'-AGT CAGCGACTCCGTCGACCAC-3'; *Cybb* (NOX2): forward 5'-ACTGCGGAGA GTTTGGAAGA-3', reverse 5'-GGTGATGACCACCTTTTGTCT-3', probe 5'-GAG GTTTGGTTCGGTTTGGCGA-3'; and *Cyba* (p22^{phox}): forward 5'-AAAGAG AAAAAGGGGTCCA-3', reverse 5'-ATGGCTGCCAGCAGATAGAT-3', probe 5'-ACTACGTCCGGGCTGCCCTC-3'.

Microarray analysis. Statistical analyses of microarray data were performed using the R programming language (<http://r-project.org>). Microarray data were normalized using the RMA method²⁷. Data were prefiltered to remove probes that were not mapped to an annotated Entrez gene. We also filtered our data to retain only a single probe per gene, selecting the probe with the highest variance, if multiple probes were found for the gene²⁸. For differential expression analysis, the limma R package was used²⁹. We modelled the synergistic regulation of gene expression by the combined IFN- γ and LPS treatment as an interaction term in our linear model. This model will identify changes that are significantly different from the sum of the individual treatments. Multiple test correction was done using the method of Benjamini and Hochberg³⁰. Genes were considered significantly different if they changed more than 1.4-fold at a false discovery rate of 0.05. Genes were further filtered for immune-cell-specific expression using the gene sets defined by the Immune Response In Silico (IRIS) project³¹. As the IRIS-defined gene sets were derived from human immune cells, we mapped the human genes to mouse orthologues using the HomoloGene database³². Genes from all IRIS-defined categories were included in the analysis. Data were submitted to the NCBI (accession number GSE53986).

Subcellular fractionation. A discontinuous sucrose gradient was used for fractionation of cells as described³³. In brief, cells were collected either by scraping (RAW cells) or by trypsinization (BMDMs), washed with PBS, re-suspended in isotonic solution (10 mM Tris, pH 7.5, 0.5 mM MgCl₂), incubated on ice for 15 min and homogenized using a Dounce homogenizer. Cell disruption was confirmed by trypan blue staining. Homogenate was made isotonic with 1.46 M sucrose (250 mM final concentration) and centrifuged at 10,000g for 5 min to remove debris and unbroken cells. The homogenate was then layered onto a sucrose gradient of 0.58 M, 0.88 M and 1.1 M, and ultracentrifuged at 100,000g for 2 h using TLS5 (Beckman Coulter optima). Five fractions numbered 1–5 were collected and analysed by western blot. ER and plasma membranes were enriched in the pellet and fraction 3, respectively. Fractions 1 and 2 were enriched in Golgi apparatus and lysosome.

Tissue lysates, immunoprecipitation and western blot analysis. For lysing tissues to analyse NRROS expression, C57BL/6J mice were perfused with 30 ml PBS, organs were collected and homogenized in RIPA buffer containing protease and phosphatase inhibitor cocktail (Halt protease and phosphatase inhibitor cocktail from Pierce). To analyse NRROS expression in different immune cell types, FACS-sorted cells were lysed in RIPA buffer containing protease and phosphatase inhibitor cocktail. For inhibitor treatments in BMDMs, cells were cultured in the presence of cycloheximide ($10 \mu\text{g ml}^{-1}$, Sigma), MG132 ($10 \mu\text{g ml}^{-1}$, Calbiochem) or chloroquine ($10 \mu\text{g ml}^{-1}$, Calbiochem) for the indicated time, collected and lysed in RIPA buffer. Neutrophils were lysed in solution containing 10% glycerol, 3% SDS, 1 mM PMSF and 5 mM NaF, and heated immediately at 95 °C for 5 min. Equal amounts of proteins, measured using BCA assay (Pierce), were separated by SDS-PAGE on 4–12% Bis-Tris gels (Invitrogen) and transferred onto nitrocellulose membranes using iBlot apparatus (Invitrogen). Membranes were blocked for 1 h at room temperature with 5% milk in TBST and probed with indicated antibody in 5% milk overnight at 4 °C.

The following primary antibodies were used: NOX2 (gp91^{phox}) (54.1, catalogue no. sc-130543), p22^{phox} (FL-195, catalogue no. sc-20781), p47^{phox} (D10, catalogue no. sc-17845), p40^{phox} (H-300, catalogue no. sc-30087), GFP (B-2, catalogue no. sc-9996) from Santa Cruz Biotechnology, p67^{phox} (rabbit polyclonal, catalogue no. 07-002) from Millipore, phospho-p40^{phox} (T154, catalogue no. 4311), β -integrin (rabbit polyclonal, catalogue no. 4706), Na-K ATPase- α (rabbit polyclonal, catalogue no. 3010), GAPDH-horseradish peroxidase (HRP) (14C10, catalogue no.

3683), HSP90 (C45G5, catalogue no. 4877), HSP70 (D69, catalogue no. 4876), CHIP (C3B6, catalogue no. 2080) from Cell Signaling Technology, phospho-p47^{phox} (Ser370) (rabbit polyclonal, catalogue no. A1171) from Assay Biotechnology, calnexin (rabbit polyclonal, catalogue no. SPA-860) from Stressgen Biotechnologies, gos28 (rabbit polyclonal, catalogue no. GTX61678) from GeneTex, actin (rabbit polyclonal, catalogue no. A5060) from Sigma-Aldrich. Anti-NRROS monoclonal antibody was generated by immunizing Armenian hamsters with *E. coli*-derived recombinant NRROS (extracellular domain). Secondary antibodies include HRP-conjugated anti-rabbit (catalogue no. 7074), anti-mouse (catalogue no. 7076) (Cell Signaling Technology) and anti-Armenian hamster (Jackson ImmunoResearch, catalogue no. 127-035-160). Image quantification was performed using ImageJ software (NIH).

For glycosidase treatment, lysates were incubated with PNGase-F (New England Biolabs) for 1 h or with Endo-H (New England Biolabs) for about 18 h according to the manufacturer's instructions. For NOX2 immunoprecipitation, lysates were incubated with 10 μg of the antibody overnight at 4 °C, immunoprecipitates were captured with protein A/G ultralink resin (Thermo Scientific) and eluted with sample buffer. For anti-Flag immunoprecipitation, lysates were incubated with EZview Red ANTI-FLAG M2 affinity gel (Sigma) at 4 °C for 4 h and eluted with 3 \times Flag peptide (Sigma) according to manufacturer's instructions.

[³⁵S]Methionine metabolic labelling. Metabolic labelling of BMDMs was performed as described^{34,35}. In brief, BMDMs were primed with IFN- γ (10 ng ml^{-1}) for 24 h. Cells were washed and incubated with $25 \mu\text{Ci ml}^{-1}$ [³⁵S]methionine (Perkin Elmer) in DMEM high glucose medium without methionine (Life Technologies) supplemented with 50 U ml⁻¹ penicillin, 50 mg ml⁻¹ streptomycin, 2 mM L-glutamine, sodium pyruvate and 10% FBS. After 1 h of labelling, cells were either collected (for time 0) or incubated further for the indicated time in complete DMEM with non-radiolabelled methionine.

Cells were washed three times with cold PBS, and then lysed in lysis buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 1 mM EDTA, 1% Triton X-100) containing protease and phosphatase inhibitor cocktail (Halt protease and phosphatase inhibitor cocktail from Pierce). Lysates were pre-cleared with control agarose beads (Pierce) for 1 h and incubated with anti-NOX2 antibody overnight. Immunoprecipitates were captured using Protein A/G ultralink resin, eluted with sample buffer and separated by SDS-PAGE. Gels were dried using a gel drier and exposed to X-ray film.

For experiments with Endo-H digestion, immunoprecipitates were eluted in Endo H buffer (100 mM Na-acetate, pH 5.6, 0.1% SDS, 1 mM PMSF)³⁶. Equal amounts of the eluted samples were incubated with or without Endo-H (New England Biolabs) for about 18 h and samples were used in SDS-PAGE as described above. Densitometry analysis was performed using ImageJ software and data were normalized to a nonspecific band in the input samples.

For experiments using MG132, cells were pre-incubated with 20 μM MG132 (Calbiochem) for 1 h before pulse. The MG132 treatment was continued during pulse and until end of the chase period.

Metabolic labelling and NOX2 immunoprecipitation in 293T cells was carried as described for BMDMs except that cells were pulsed with $100 \mu\text{Ci ml}^{-1}$ [³⁵S]methionine for 15 min instead of 1 h. Overexpressing NOX2 by transient transfection enabled us to pulse cells for a shorter duration (for example, 15 min) and still be able to immunoprecipitate and detect NOX2.

RAC1 activation assay. The activity of RAC1 in wild-type and NRROS-deficient BMDMs was measured using RAC1 activation assay kit (Cell Biolabs) according to the manufacturer's instructions. In brief, wild-type and NRROS-deficient BMDMs were primed with 10 ng ml^{-1} IFN- γ for 24 h. Cells were stimulated with $100 \mu\text{g ml}^{-1}$ zymosan for 5, 15 or 30 min. Cell lysates were prepared and used immediately for RAC1 pulldown with the p21-activated protein kinase (PAK)-p21 binding domain (PBD) beads. The activated RAC1 was detected by anti-RAC1 antibody supplied with the kit.

Immunofluorescence staining. RAW 264.7 cells stably expressing Flag-NRROS were grown overnight in chamber slides. Cells were then fixed with 4% paraformaldehyde, washed with PBS, and treated with 50 mM NH₄Cl/PBS. After rinsing with PBS, cells were permeabilized and blocked in saponin buffer (0.4% Saponin, 1% BSA, 2% normal goat serum in PBS). Mouse anti-Flag (Sigma-Aldrich), rabbit anti-calnexin (Enzo Life Sciences) and rat anti-LAMP1 (Santa Cruz Biotechnology) antibodies were used. Secondary goat anti-mouse Alexa 647 was purchased from Invitrogen and goat anti-rabbit-Cy3 and donkey anti-rat-Cy5 antibodies were from Jackson ImmunoResearch.

Confocal microscopy. Images were captured by a Leica SPE laser scanning confocal microscope (Leica) equipped with Leica DM5500Q under $63\times 1.30 \text{ NA CS ACS APO}$ oil-immersion objective controlled by LAS AF image acquisition and processing software. Excitation and emission for DAPI, Alexa 488, Cy3 and Cy5 were used at default settings set by the manufacturer (Leica).

Phagocytosis assay. dsRed *E. coli* (gift from K. Huynh, Genentech) were opsonized with fresh mouse serum, and BMDMs primed with 10 ng ml^{-1} IFN- γ for

24 h were allowed to phagocytose the bacteria for 30 min. Cells were collected and analysed by FACS.

In vitro bacterial killing assay. Macrophages plated at 0.25×10^6 cells per well of a 24-well plate were primed with 10 ng ml^{-1} of IFN- γ for 24 h in antibiotics-free complete DMEM. Cells were infected with serum opsonized *E. coli* strain 26 at m.o.i. of 50 in 250 μl total volume by centrifugation at room temperature for 5 min at 1,000 r.p.m. to synchronize the infection and further incubated at 37°C for 30 min. Medium was replaced with $100 \mu\text{g ml}^{-1}$ of gentamicin (Invitrogen)-containing medium. After an additional 5 min (considered as 30-min time point), 30 min (60-min time point) and 90 min (120-min time point), cells were washed twice with 1 ml PBS and lysed in 250 μl of 0.1% Triton X-100 in PBS. Twofold serial dilutions of the lysates were prepared in PBS and 5 μl was dropped on a LB agar plate with no antibiotics. Colonies were counted after 18–24 h incubation at 37°C . Total protein in the lysates was measured using BCA assay and the colony numbers were normalized to the amount of protein.

Neutrophil bacterial killing assay was performed as described previously³⁷. In brief, twofold dilutions ($2.5\text{--}0.625 \times 10^6$) of neutrophils were added to 25×10^6 c.f.u. *E. coli* strain 26 and fresh mouse serum (5% final concentration) in a total volume of 1 ml, in triplicate. The tubes were rotated for 2 h and 100 μl of the sample was diluted with 1 ml sterile distilled water (pH adjusted to 11 by NaOH) to lyse cells. Serial dilutions were prepared and plated on LB agar plates without antibiotics. Bacterial colonies were counted after 24 h incubation at 37°C .

In vivo bacterial challenge. For *L. monocytogenes* infection studies, age- and sex-matched (6–8 weeks old) wild-type, NRROS-deficient, NOX2-deficient or NRROS NOX2 double-deficient littermate mice were administered intravenously with *L. monocytogenes* (ATCC 43251) (0.1×10^6 c.f.u. for females and 0.5×10^6 c.f.u. for males), and were monitored for 14 days for morbidity and mortality. For bacterial burden experiments, mice were administered intravenously with 1×10^6 c.f.u. *L. monocytogenes* or 2×10^8 c.f.u. *E. coli* strain 26, and killed at 48 h (*L. monocytogenes*) or 6 h (*E. coli*). Spleens and livers were collected in 50-ml tissue grinder tubes (VWR International) and homogenized in 0.1% Triton X-100 in PBS. Twofold serial dilutions of the homogenates were prepared and plated in triplicate on brain heart infusion agar. Bacterial colonies were counted after 24 h. Total protein in the lysate was measured using BCA assay and colony numbers were normalized to the amount of protein.

For ROS neutralization, mice were injected intravenously with either a cocktail of scavengers³⁸ that included 15 mg kg^{-1} EUK134 (Cayman chemicals) 20 mg kg^{-1} FeTPPS (5,10,15,20-tetrakis(4-sulphonatophenyl)porphyrinato iron (III) chloride) (Calbiochem) and 50 mg kg^{-1} PBN (*N*-tert-butyl- α -phenylnitron) (Sigma-Aldrich) or 2.5% DMSO (Sigma-Aldrich) as vehicle control in a total volume of 200 μl at 60 min before administering *L. monocytogenes*. The spleens and livers were collected after 24 h and processed as described above, for enumeration of bacterial burden.

For all *L. monocytogenes* infection experiments, a sample size of at least five mice per genotype was used in each independent experiment except for data in Fig. 2i, j where three mice were used for some groups. Mice of both genders were used. No animals were excluded from any study. No blinding or randomization method was used to group the animals.

Induction of EAE. EAE was induced as described previously³⁹. In brief, age- and sex-matched littermate mice were immunized subcutaneously at the base of the tail with 300 μg MOG_{35–55} peptide emulsified in 800 μg *Mycobacterium tuberculosis*-supplemented CFA (Difco Laboratories). On day 0 and day 2 post-immunization, 200 ng pertussis toxin (List Biologicals) was administered intraperitoneally. Clinical disease was scored daily starting on day 8 until day 30 as follows: 0, no clinical disease; 1, limp tail or hind limb weakness but not both; 2, limp tail and hind limb weakness; 3, partial hind limb paralysis; 4, complete hind limb paralysis; 5, moribund/death.

For ROS neutralization, a cocktail of scavengers was used as described previously³⁸. In brief, mice were administered intraperitoneally every 12 h with the aforementioned cocktail or vehicle control starting from day 9 or day 10 post-immunization depending on the appearance of first clinical signs.

All the mice used for EAE experiments shown in Fig. 3 were bone marrow chimaeric mice. The chimaeric mice were generated when they were around 6 weeks

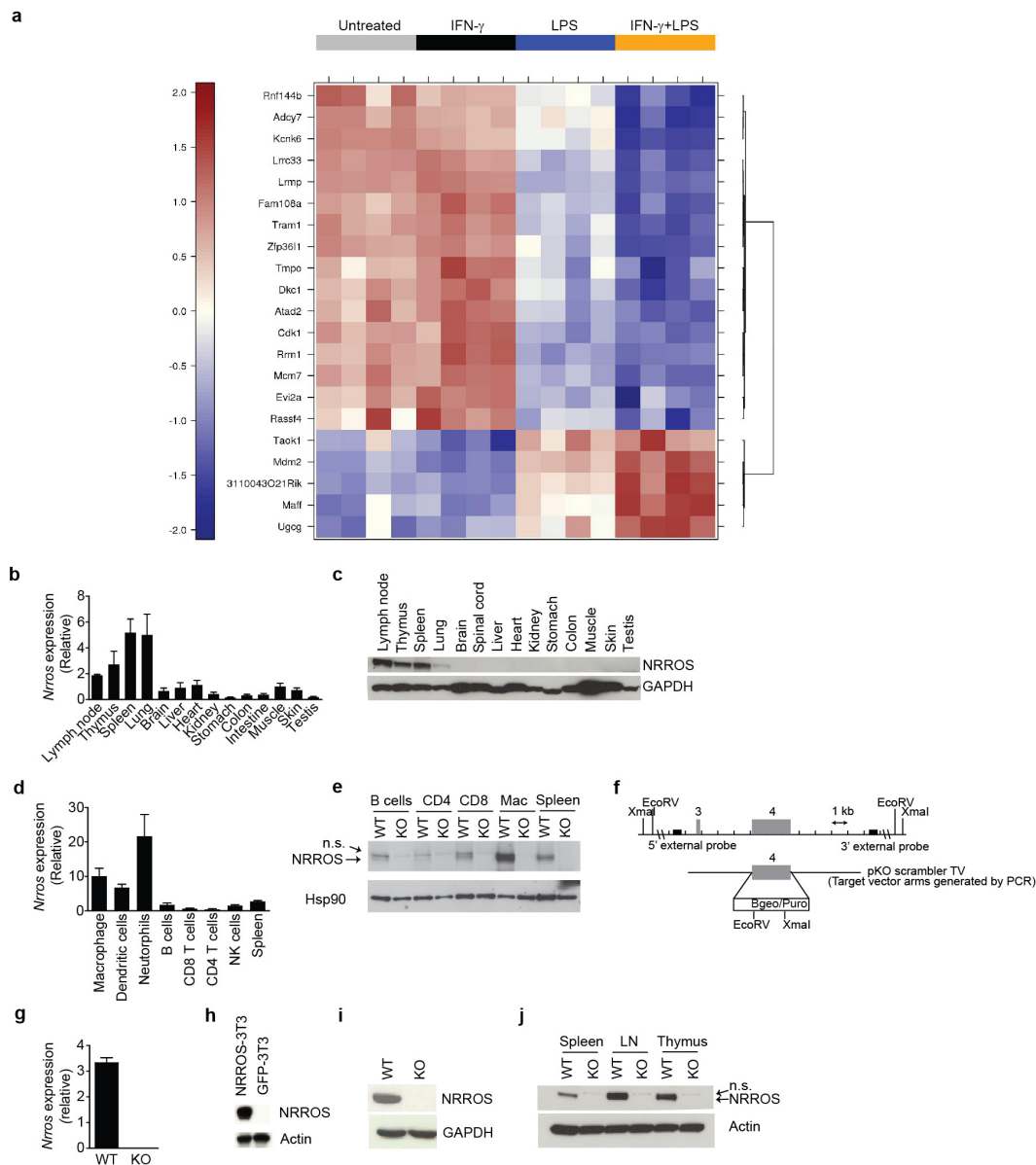
old and 8–12 weeks, or even longer, was allowed for reconstitution. Therefore, mice were 14–20 weeks old at the time of immunization. Male mice were used for most of the experiments. *Nrros*^{fl/fl} LysM-Cre^{pos} and *Nrros*^{fl/fl} LysM-Cre^{neg} mice used for data shown in Extended Data Fig. 5 were 6–8 weeks old.

A sample size of at least five mice per genotype was used in each independent experiment. No animals were excluded from any study. No blinding or randomization method was used in grouping the mice.

Measurement of oxidative damage in CNS tissues. Oxidative damage in the CNS of EAE-induced mice was assessed by measuring malondialdehyde levels⁴⁰. Brain and spinal cords from EAE-induced mice were collected and homogenized in PBS. Protein estimation was done by BCA assay. Homogenates were used to measure the levels of MDA by ELISA using OxiSelect MDA adduct ELISA kit (Cell Biolabs) or by western blot using OxiSelect MDA immunoblot kit (Cell Biolabs) according to the manufacturer's recommendations.

Statistical analysis. Statistical significance was calculated by Prism 5 software (Graphpad) using appropriate statistical test mentioned in the figure legends. For ROS measurements by luminol, area under curve (AUC) was calculated for each individually stimulated sample (at least three samples per group). The AUC values were then used in unpaired *t*-test with Welch's corrections to generate a *P* value for each experiment. The average AUCs from different experiments were used in paired *t*-test assuming equal variance to generate *P* values when comparing multiple experiments. Any other appropriate statistical methods applied for a data set are described in the figure legends. A *P* value of less than 0.05 was considered as significant.

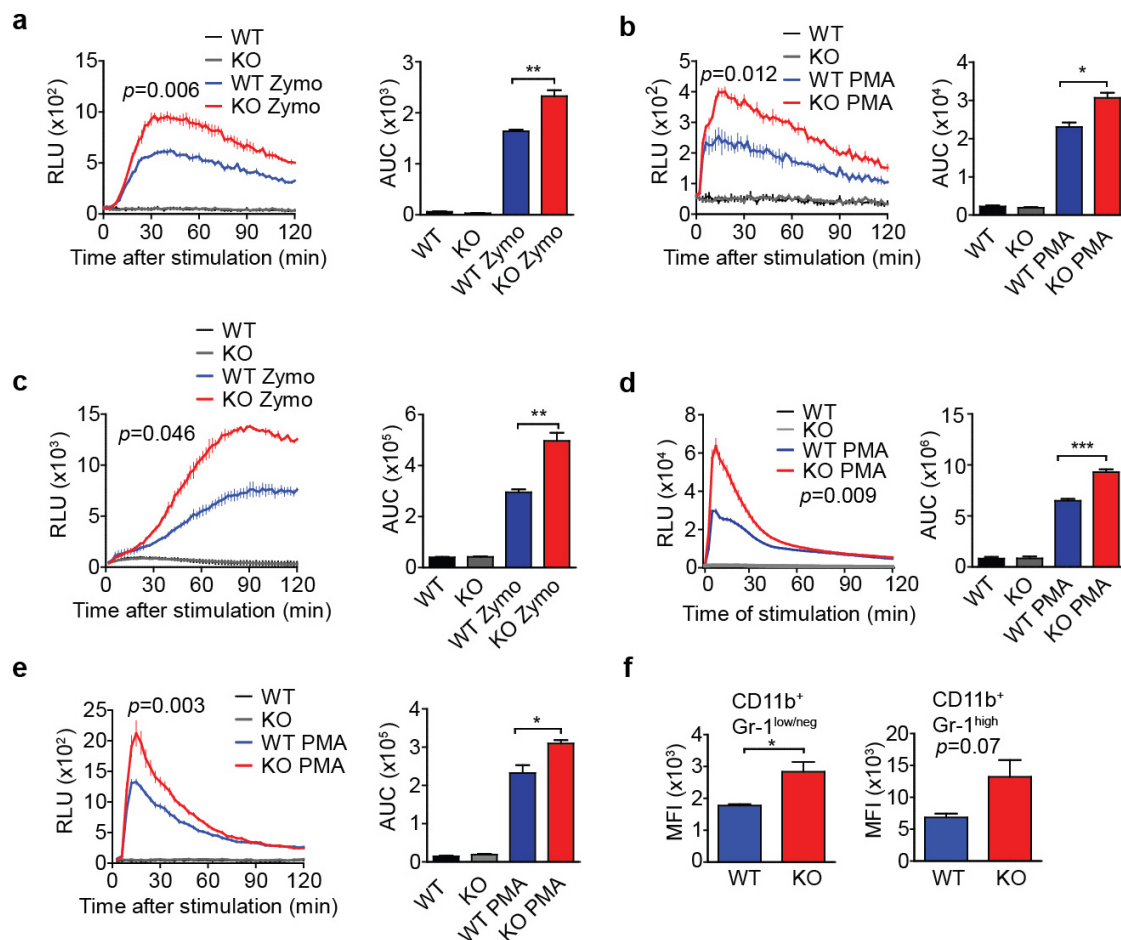
25. Tang, T. et al. A mouse knockout library for secreted and transmembrane proteins. *Nature Biotechnol.* **28**, 749–755 (2010).
26. Mantei, A. et al. siRNA stabilization prolongs gene knockdown in primary T lymphocytes. *Eur. J. Immunol.* **38**, 2616–2625 (2008).
27. Irizarry, R. A. et al. Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**, 249–264 (2003).
28. Bourgon, R., Gentleman, R. & Huber, W. Independent filtering increases detection power for high-throughput experiments. *Proc. Natl Acad. Sci. USA* **107**, 9546–9551 (2010).
29. Smyth, G. K. Linear models and empirical bayes methods for assessing differential expression in microarray experiments. *Stat. Appl. Genet. Mol. Biol.* **3**, Article3 (2004).
30. Benjamini, Y. & Hochberg, Y. Controlling the false discovery rate: a practical and powerful approach to multiple testing. **57**, 289–300 *J. Roy. Stat. Soc. B* (1995).
31. Abbas, A. R. et al. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun.* **6**, 319–331 (2005).
32. Sayers, E. W. et al. Database resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **40**, D13–D25 (2012).
33. Li, Y. et al. Enrichment of endoplasmic reticulum with cholesterol inhibits sarcoplasmic-endoplasmic reticulum calcium ATPase-2b activity in parallel with increased order of membrane lipids: implications for depletion of endoplasmic reticulum calcium stores and apoptosis in cholesterol-loaded macrophages. *J. Biol. Chem.* **279**, 37030–37039 (2004).
34. DeLeo, F. R. et al. Processing and maturation of flavocytochrome b558 include incorporation of heme as a prerequisite for heterodimer assembly. *J. Biol. Chem.* **275**, 13986–13993 (2000).
35. Barry, K. C., Fontana, M. F., Portman, J. L., Dugan, A. S. & Vance, R. E. IL-1 α signaling initiates the inflammatory response to virulent *Legionella pneumophila* in vivo. *J. Immunol.* **190**, 6329–6339 (2013).
36. Green, S. A., Zimmer, K. P., Griffiths, G. & Mellman, I. Kinetics of intracellular transport and sorting of lysosomal membrane and plasma membrane proteins. *J. Cell Biol.* **105**, 1227–1240 (1987).
37. Czubynski, C. J., Canono, B. P., Henson, P. M. & Campbell, P. A. Genetically determined resistance to listeriosis is associated with increased accumulation of inflammatory neutrophils and macrophages which have enhanced listericidal activity. *Immunology* **55**, 511–518 (1985).
38. Nikić, I. et al. A reversible form of axon damage in experimental autoimmune encephalomyelitis and multiple sclerosis. *Nature Med.* **17**, 495–499 (2011).
39. Hu, Y. et al. IL-17RC is required for IL-17A- and IL-17F-dependent signaling and the pathogenesis of experimental autoimmune encephalomyelitis. *J. Immunol.* **184**, 4307–4316 (2010).
40. Barabuti, N. & Schally, A. V. Antioxidant activity of growth hormone-releasing hormone antagonists in LNCaP human prostate cancer line. *Proc. Natl Acad. Sci. USA* **105**, 20470–20475 (2008).



Extended Data Figure 1 | Murine NRROS is primarily expressed in immune tissues, especially in phagocytes.

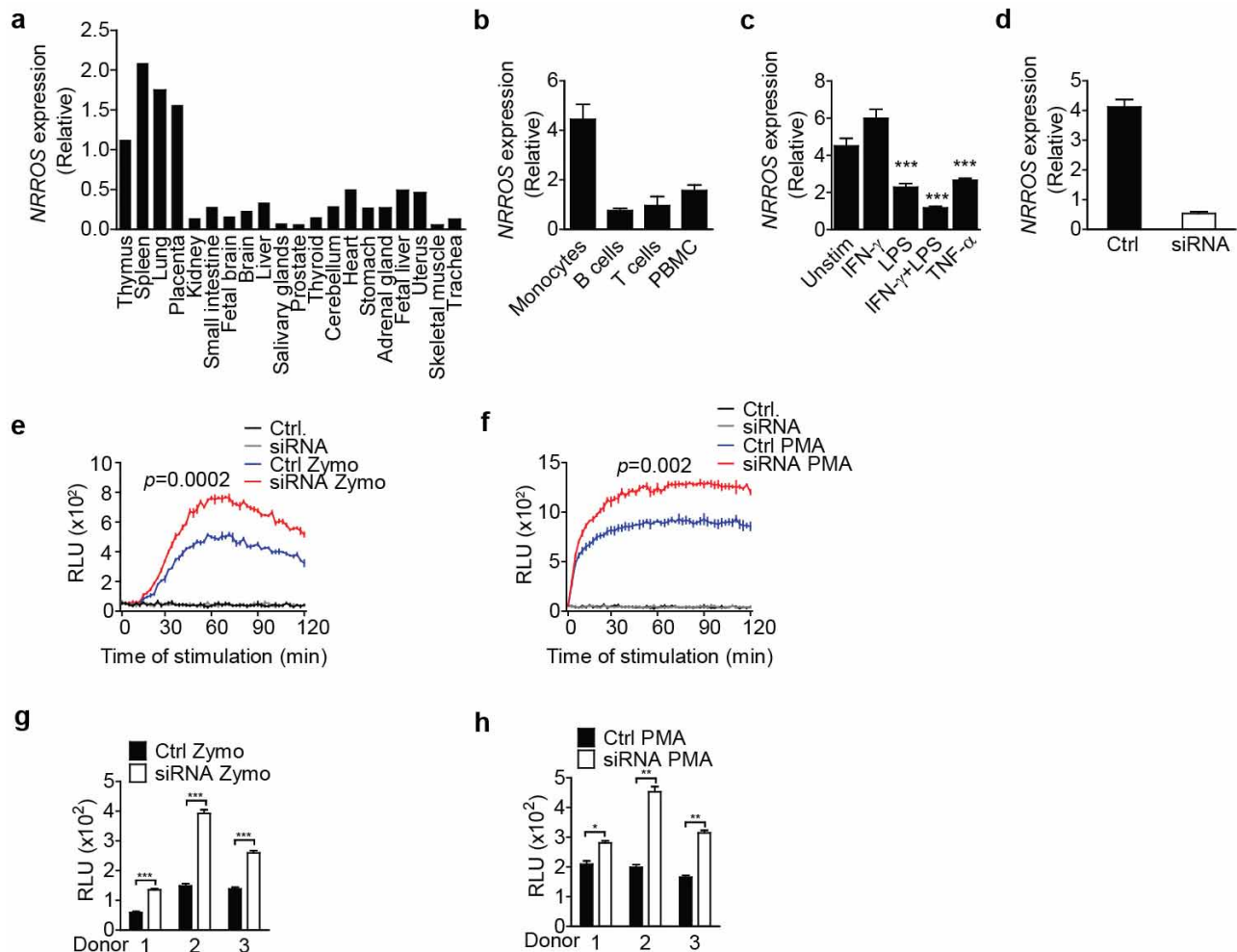
a, BMDMs were either untreated or treated with IFN- γ , LPS or both. Heat map displays genes that were preferentially expressed in leukocytes and differentially regulated by the synergistic effects of both the treatments compared to either treatment alone. **b**, Analysis of *Nrros* expression in total RNA from different mouse tissues. Expression levels were normalized to *Rpl19* ($n = 4$). **c**, Western blot analysis of NRROS protein expression in different mouse tissues. GAPDH used as loading control. **d**, Analysis of *Nrros* expression in total RNA from immune cells sorted from mouse spleen, normalized to *Rpl19* ($n = 3-5$). **e**, Western blot analysis of NRROS protein expression in the indicated cells. HSP90 used as loading control. Mac, peritoneal macrophage; n.s., non-specific band observed only in lymphoid cells but not in macrophages with anti-NRROS antibody.

f, Schematic of the strategy used to generate NRROS-deficient mice. **g**, Analysis of *Nrros* expression in total RNA from spleens of WT and KO mice ($n = 5$), normalized to *Rpl19*. **h**, Western blot analysis of NRROS protein expression in 3T3 cells overexpressing NRROS (NRROS-3T3) and control cells (GFP-3T3) to screen anti-NRROS antibody. Actin used as loading control. **i**, Western blot analysis of NRROS protein expression in WT and KO BMDMs. GAPDH as loading control. **j**, Western blot analysis of NRROS protein expression in immune tissues from WT and KO mice. Actin used as loading control. Although there was no detectable specific band in KO BMDMs, a weak, nonspecific band was observed in lymphoid cells from both WT and KO mice that was present just above the specific NRROS band. Error bars, s.e.m. Data in **c**, **e** and **j** are representative of at least three independent experiments.



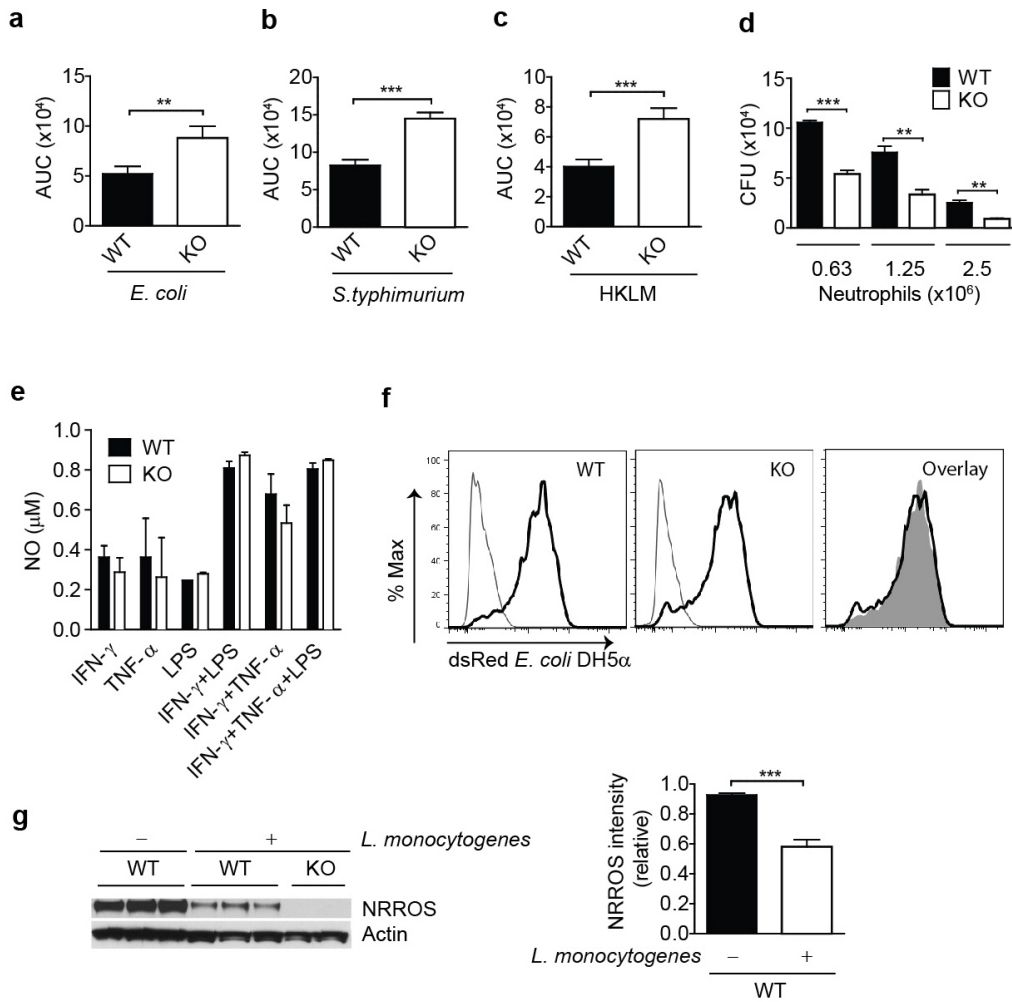
Extended Data Figure 2 | NRROS negatively regulates ROS production in various myeloid subsets. **a–e**, ROS production in thioglycollate-elicited peritoneal macrophages (**a**, **b**), neutrophils (**c**, **d**) and BMDM (**e**) stimulated with zymosan (zymo) (**a**, **c**) and phorbol 12-myristate 13-acetate (PMA) (**b**, **d**, **e**). Cells in **a**, **b**, **e** were primed with IFN- γ before ROS induction. Left panels, ROS kinetic plots of one representative experiment with three independent samples per group. Right panels, AUC from at least three

independent experiments. **f**, ROS production in splenocytes from WT and KO mice ($n = 3$) stimulated with PMA in the presence of CM-H2DCFDA and analysed by FACS. MFI of CD11b⁺ Gr-1^{low/neg} cells (left) and CD11b⁺ Gr-1^{high} (right) are shown. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Unpaired Student's t -test with Welch's corrections (**a–e**, left), paired Student's t -test (**a–e**, right), unpaired Student's t -test (**f**). Data in **f** are representative of two independent experiments.



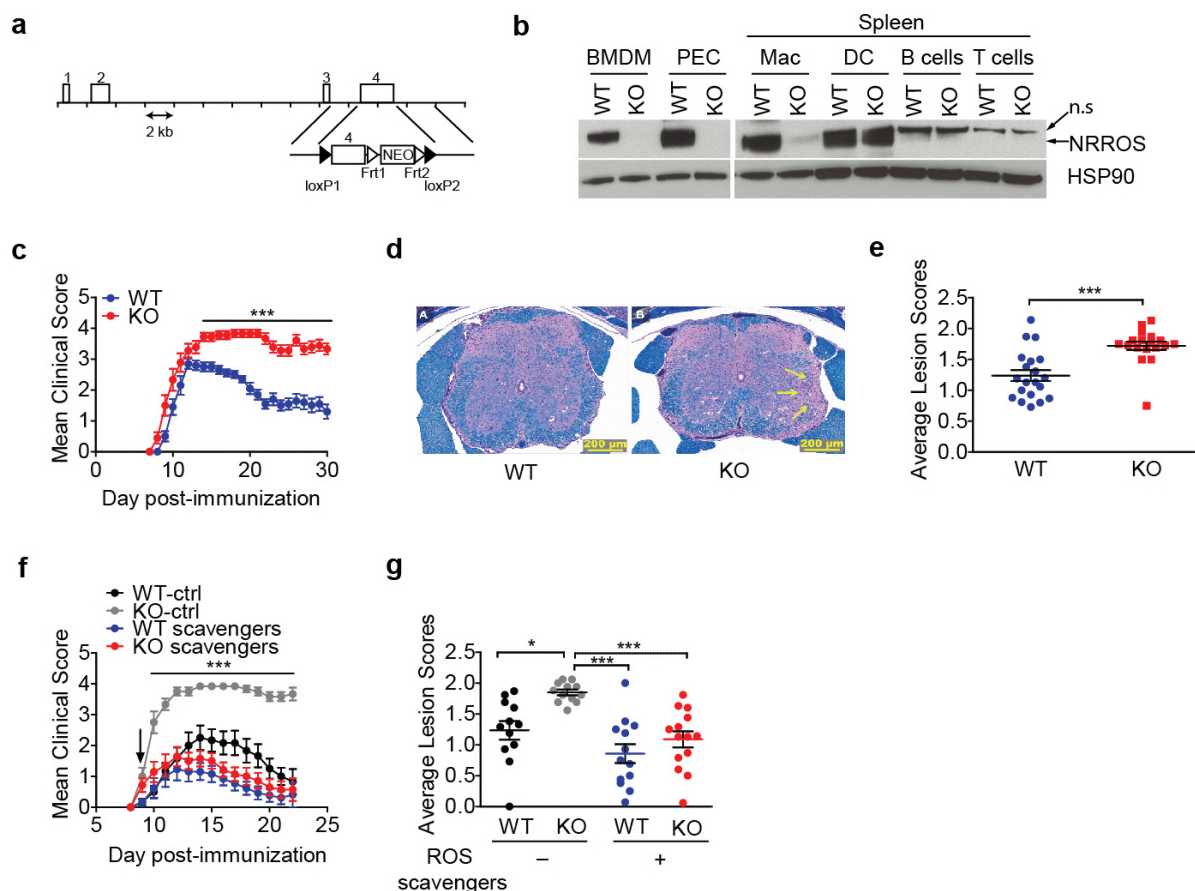
Extended Data Figure 3 | NRROS negatively regulates ROS production in human monocyte-derived macrophages (MDMs). **a, b**, Analysis of *NRROS* expression in total mRNA from human tissues (**a**) and cells purified from healthy donors (**b**) ($n = 3$). **c**, *NRROS* expression in MDMs stimulated as indicated ($n = 3$). **d**, *NRROS* expression in MDMs treated with non-targeting (Ctrl) or *NRROS* siRNA (siRNA). **e, f**, Kinetic graph of ROS production from cells shown in **d**. Cells were primed with IFN- γ and were then either

unstimulated or stimulated with zymosan (**e**) or PMA (**f**). **g, h**, ROS production by IFN- γ -primed MDMs stimulated, in triplicate, with zymosan (**g**) and PMA (**h**) from three individual donors. Values at the peak production (60 min) are shown. Data in **a–d** are normalized to endogenous control *RPL19*. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$, *** $P < 0.001$. Unpaired Student's *t*-test with (**e, f**) and without (**g, h**) Welch's corrections. Data are representative of two independent experiments with three donors in each experiment.



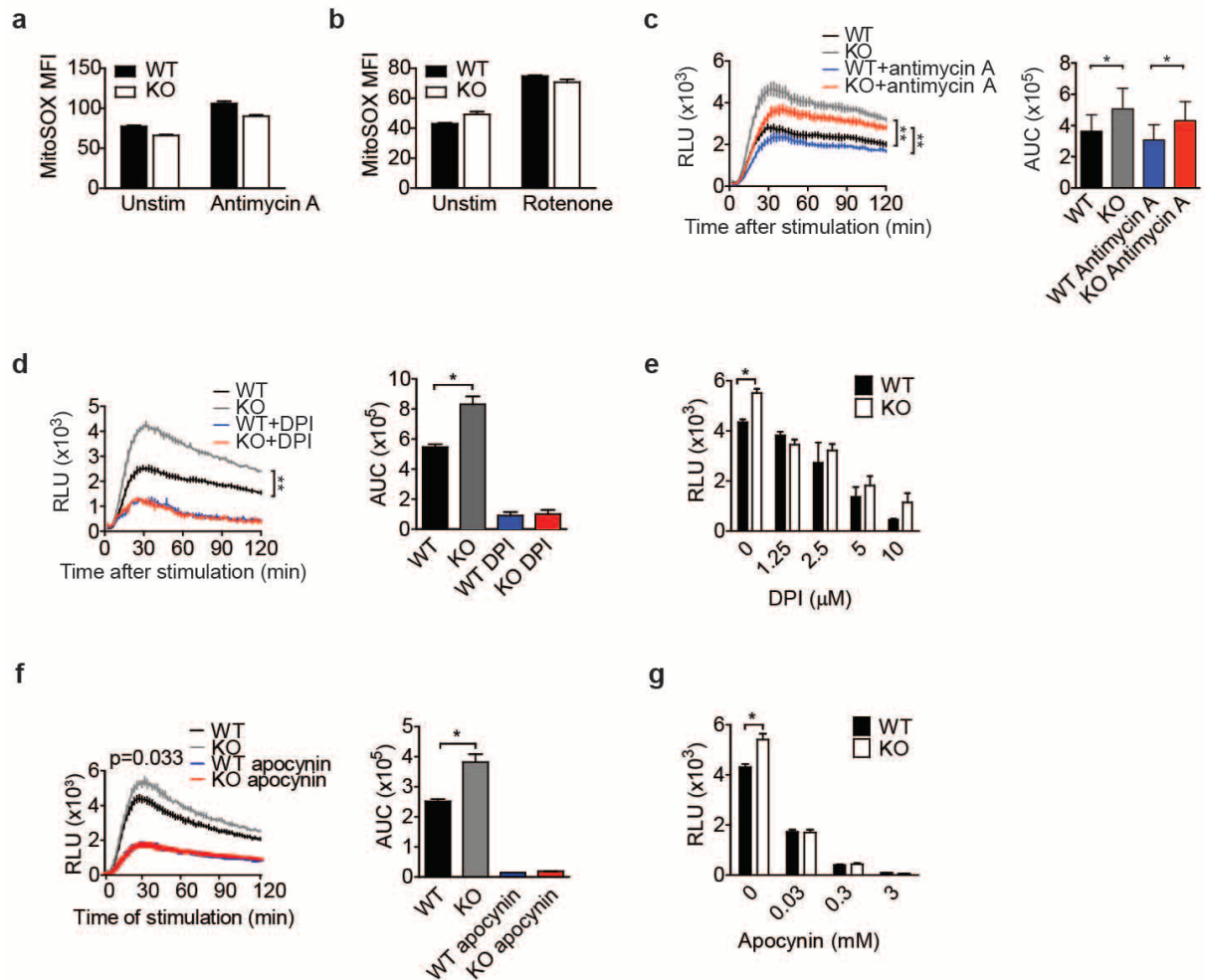
Extended Data Figure 4 | NRROS regulates ROS production and bactericidal activities of phagocytes but not phagocytosis or nitric oxide production by them. **a–c**, ROS production in IFN- γ -primed BMDMs stimulated with *E. coli* (**a**), *S. typhimurium* (**b**) and heat-killed *L. monocytogenes* (HKLM) (**c**). Averaged AUC from at least four independent experiments is plotted as bar graphs. **d**, *In vitro* *E. coli* killing by neutrophils purified from bone marrow of either WT or KO mice ($n = 3$). **e**, Nitric oxide (NO) in culture supernatant of BMDMs stimulated as indicated ($n = 3$). **f**, Representative

histograms showing phagocytosis of serum-opsonized dsRed-*E. coli* by BMDMs. Thin line, untreated cells; thick line, treated cells. In the overlay histograms on the right: filled area, WT cells; thick line, NRROS-deficient cells. **g**, Left, NRROS expression in CD11b $^+$ cells purified from spleens of uninfected WT and *L. monocytogenes*-infected WT and KO mice. Actin used as loading control. Right, densitometry analyses relative to actin. Error bars, s.e.m. *** $P < 0.001$. Paired (**a–c**) or unpaired (**d, g**) Student's *t*-test. Data in **d–g** are representative of at least two independent experiments.



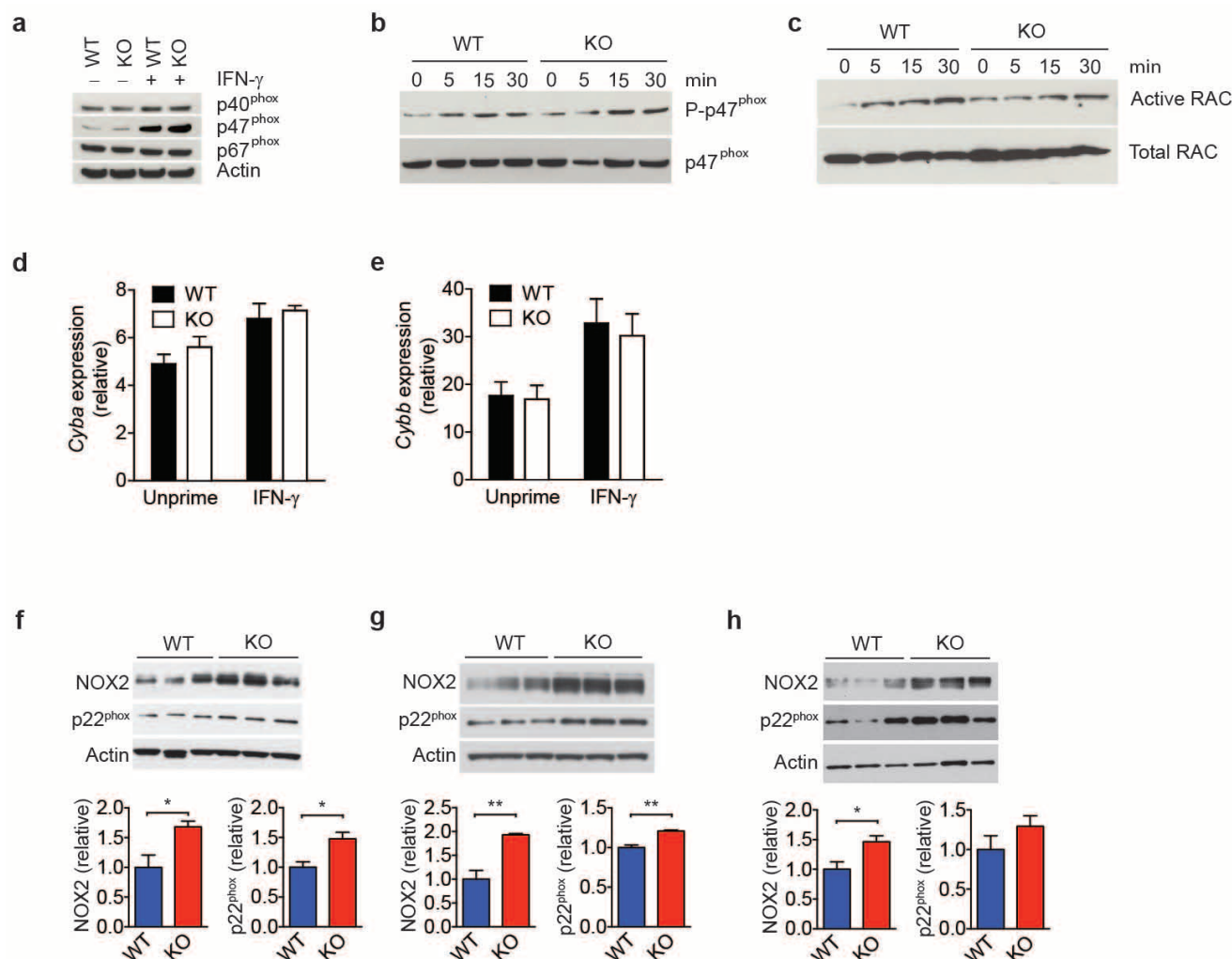
Extended Data Figure 5 | Myeloid-specific deletion of NRROS leads to exacerbated EAE. **a**, Schematic of the strategy used to generate *Nrros*^{fl/fl} mice. Numbers represent exons in *Nrros* gene. WT, *Nrros*^{fl/fl} LysM-Cre^{neg}; KO, *Nrros*^{fl/fl} LysM-Cre^{pos}. **b**, Western blot analysis of NRROS protein in immune cells from WT and KO mice. Actin used as loading control. Mac, macrophage; PEC, peritoneal macrophage. **c–e**, Clinical course (c), representative images of the spinal cord haematoxylin and eosin staining (d), and average histopathological lesion scores (day 30) (e) of EAE-induced WT (*n* = 20) and KO (*n* = 18) mice. Arrows in d indicate extensive myelinopathy of the lateral

white matter tracts. Luxol fast blue staining of myelin is in blue. **f**, **g**, Clinical course (f) and lesion severity as assessed by histopathology on day 30 (g) of EAE-induced WT and KO mice administered with either vehicle control (ctrl) (*n* = 12) or a cocktail of ROS scavengers (*n* = 13 for WT and *n* = 14 for KO) every 12 h starting from day 9 (arrow) for the rest of the study. Error bars, s.e.m. **P* < 0.05, ****P* < 0.001. Two-way (c, f) or one-way (g) analysis of variance followed by Bonferroni's post-hoc analysis, unpaired Student's *t*-test (e). Data are representative of at least three (c–e) and two (b) independent experiments. Data in f, g are combined from two independent experiments.



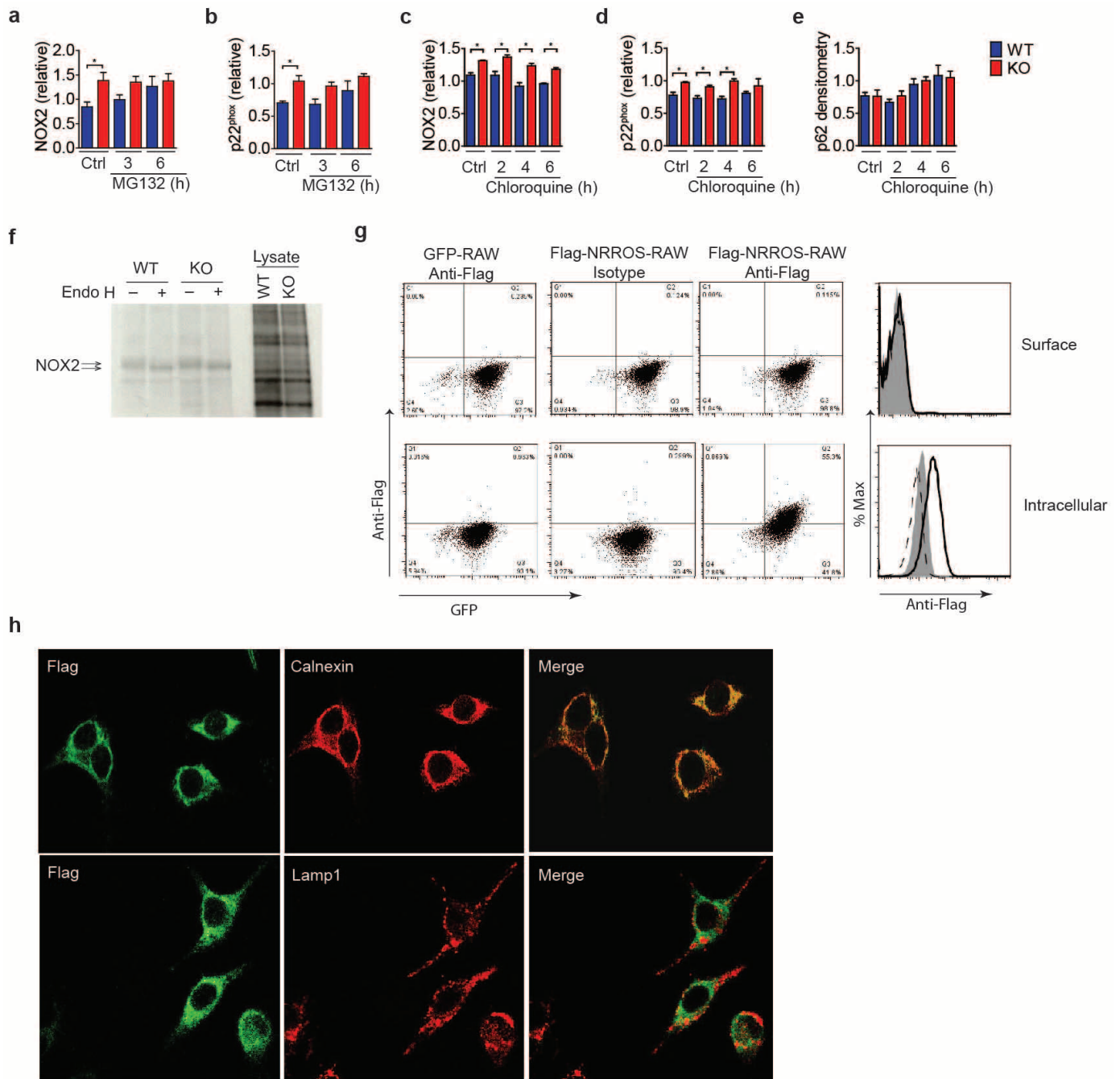
Extended Data Figure 6 | NRROS regulates NOX2-mediated ROS but not mROS generation. **a, b**, mROS production by IFN- γ -primed BMDMs from WT and NRROS-deficient (KO) mice stimulated with antimycin A (**a**) or rotenone (**b**) in the presence of MitoSOX. **c–g**, IFN- γ -primed BMDMs were pre-treated for 1 h with antimycin A (**c**), DPI (**d, e**) or apocynin (**f, g**) before ROS induction by zymosan. (**c, d, f**). Left panels, ROS kinetic plots of a representative experiment with three independent samples per group; right

panels, averaged AUC from three independent experiments. **e, g**, Dose response of DPI (**e**) or apocynin (**g**) treatment in WT and KO cells. Values at the peak of the curve are used for the bar graph. Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$. Unpaired Student's *t*-test with Welch's corrections (**c, d, f**, left), paired Student's *t*-test (**c, d, f**, right), unpaired Student's *t*-test (**e, g**). Data are representative of two (**a, b, e, g**) or three (**c, d, f**) independent experiments.



Extended Data Figure 7 | No difference between WT and NRROS-deficient BMDMs in expression and activation of the cytosolic members of NADPH oxidase complex. **a**, Western blot analysis of p40^{phox}, p47^{phox} and p67^{phox} in the lysates of BMDMs from WT and NRROS-deficient (KO) mice either unprimed (-) or primed (+) with IFN- γ . Actin used as loading control. **b**, **c**, phospho-p47^{phox} (P-p47^{phox}) and total p47^{phox} (**b**), active Rac and total Rac (**c**) in IFN- γ -primed BMDMs stimulated with zymosan for the indicated time. Active Rac was immunoprecipitated and blotted. Total Rac in the lysates used as control. **d**, **e**, Taqman analysis of *Cyba* (p22^{phox}) (**d**) and *Cybb* (NOX2)

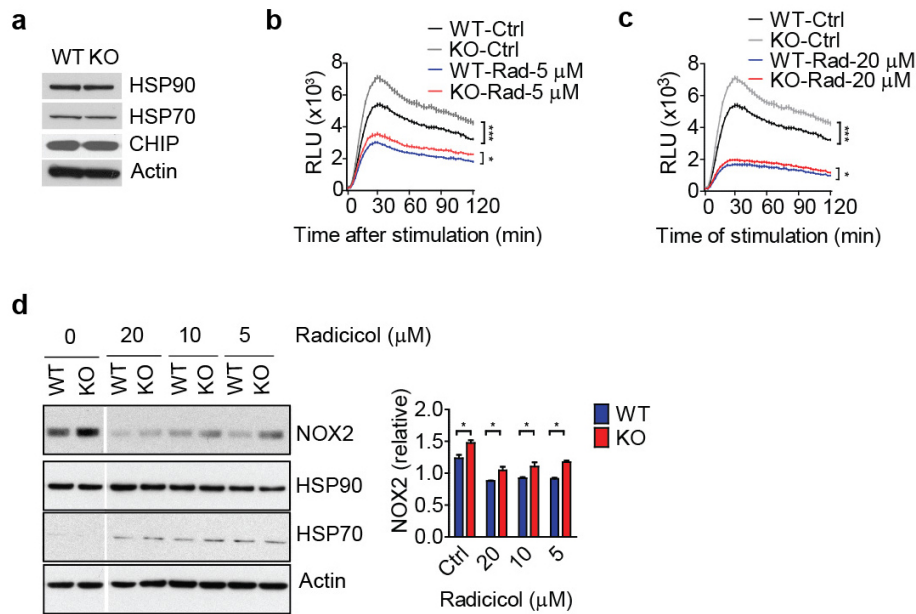
(**e**) expression in total RNA from BMDMs either primed with IFN- γ or left alone (Unprime). Expression levels were normalized to endogenous control, ribosomal protein L19 (*Rpl19*) ($n = 3$). **f**–**h**, Top panels, NOX2 and p22^{phox} expression in thioglycollate-elicited peritoneal macrophages (**f**), in splenocytes from mice infected with *L. monocytogenes* for 24 h (**g**) and in total CNS cells from EAE-induced mice on day 15 post-immunization (**h**). Bottom panels, densitometry analyses relative to actin ($n = 3$). Error bars, s.e.m. * $P < 0.05$, ** $P < 0.01$ (unpaired Student's *t*-test). Data are representative of three (**a**–**c**) and two (**f**–**h**) independent experiments.



Extended Data Figure 8 | NRROS is an ER protein and regulates proteasome-mediated degradation of NOX2 and p22^{phox} proteins.

a–e, Densitometry analyses relative to actin of NOX2 (**a**, **c**), p22^{phox} (**b**, **d**) and p62 (**e**) in IFN- γ -primed BMDMs from WT and KO cells treated with MG132 (**a**, **b**) or chloroquine (**c–e**) for the indicated period ($n = 3$). p62 is shown as positive control for chloroquine treatment. **f**, Autoradiography showing immunoprecipitated NOX2 from BMDMs metabolically labelled with [³⁵S]methionine for 1 h. The immunoprecipitates were subjected to Endo H treatment for 18 h. Total proteins in the flow-through lysates is shown as loading control. **g**, RAW cells retrovirally transduced to stably express GFP (GFP-RAW) or Flag-NRROS (Flag-NRROS-RAW) were stained with either isotype control (second column) or anti-Flag antibody (first and third

columns). Cells were either non-permeabilized (top panels) or permeabilized (bottom panels). Overlay of the three populations is shown on the right. Thick line, Flag-NRROS-RAW cells stained with anti-Flag; filled area, GFP-RAW cells stained with anti-Flag; dashed line, Flag-NRROS-RAW cells stained with isotype control. **h**, RAW cells stably expressing Flag-tagged NRROS were stained with anti-Flag antibody (Flag) (green) and one of the indicated organelle markers (red), calnexin (top row) and Lamp1 (bottom row) and imaged by confocal microscopy. Colocalization of the two colours is depicted by yellow colour in the merged images on the right column. Error bars, s.e.m. **P* < 0.05 (unpaired Student's *t*-test). Data are representative of at least three independent experiments.



Extended Data Figure 9 | NRROS does not regulate NOX2 through the HSP90–HSP70–CHIP pathway. **a**, Western blot analysis of HSP90, HSP70 and CHIP in total lysates from IFN- γ -primed WT and KO BMDMs. Actin used as loading control. **b**, **c**, Zymosan-induced ROS production by IFN- γ -primed BMDMs treated with the HSP90 inhibitor radicicol at 5 μ M (**b**) and 20 μ M (**c**). A representative plot of kinetics of ROS production (three independently stimulated samples per group) from three independent experiments is shown.

d, Left, NOX2, HSP90 and HSP70 expression in IFN- γ -primed BMDMs treated with the indicated dose of radicicol. Actin used as loading control. Right, densitometry analyses relative to actin from three experiments. Error bars, s.e.m. * $P < 0.05$, *** $P < 0.001$. Unpaired Student's t -test with (**b**, **c**) and without (**d**) Welch's corrections. Data in **a** are representative of three experiments.

Extended Data Table 1 | Analysis of immune cell subsets in lymph node, spleen, thymus and blood of WT and KO mice

Lymph node	WT (n=5)	KO (n=5)
CD4 ⁺	64.1 ± 1.00	65.8 ± 0.62
CD8 ⁺	25.6 ± 0.67	24.5 ± 0.32
B220 ⁺	14.9 ± 0.86	15.6 ± 1.14
CD4 ⁺ CD25 ⁺	2.5 ± 0.10	2.7 ± 0.13

Spleen	WT (n=5)	KO (n=5)
CD4 ⁺	43.0 ± 0.98	45.1 ± 0.66
CD8 ⁺	24.0 ± 0.68	21.0 ± 0.86
B220 ⁺	42.9 ± 0.40	42.9 ± 1.85
γδ T cells	0.5 ± 0.05	0.5 ± 0.05
DX5 ⁺ (NK cells)	4.0 ± 0.25	4.02 ± 0.13
CD11b ⁺ F4/80 ⁺	4.1 ± 0.22	4.3 ± 0.07
CD11b ⁺ CD11c ⁺	0.5 ± 0.39	0.5 ± 0.03
CD11b ⁺ Gr-1 ⁺	3.2 ± 0.34	3.3 ± 0.11

Thymus	WT (n=5)	KO (n=5)
CD4 ⁺ CD8 ⁺	84.0 ± 1.14	84.3 ± 0.92
CD4 ⁺ CD8 ⁻	6.7 ± 0.78	7.1 ± 0.59
CD4 ⁻ CD8 ⁻	4.6 ± 0.43	4.2 ± 0.11
CD4 ⁻ CD8 ⁺	4.4 ± 0.38	4.8 ± 0.44

Complete Blood Count (CBC)	WT (n=5)	KO (n=5)
White blood cells	9.7 ± 1.67	10.0 ± 1.19
Neutrophils	9.7 ± 1.82	6.1 ± 0.95
Lymphocytes	81.3 ± 1.71	81.6 ± 2.87
Monocytes	6.5 ± 0.75	11.0 ± 3.02
Eosinophils	1.1 ± 0.29	0.6 ± 0.35
Basophils	1.4 ± 0.17	0.7 ± 0.44

Single-cell suspensions of lymph node, spleen and thymus were stained with the indicated surface marker and analysed by FACS. Percentages of cells in the total population are shown. Haematological analysis of the blood (complete blood count) showing the percentage of total white blood cells and the percentages of different leukocyte populations therein is shown. All the values are average ± s.e.m.

Endosomes are specialized platforms for bacterial sensing and NOD2 signalling

Norihiro Nakamura¹, Jennie R. Lill¹, Qui Phung¹, Zhaoshi Jiang¹, Corey Bakalarski¹, Ann de Mazière², Judith Klumperman², Megan Schlatter¹, Lélia Delamarre¹ & Ira Mellman¹

The detection of microbial pathogens involves the recognition of conserved microbial components by host cell sensors such as Toll-like receptors (TLRs) and NOD-like receptors (NLRs). TLRs are membrane receptors that survey the extracellular environment for microbial infections, whereas NLRs are cytosolic complexes that detect microbial products that reach the cytosol. Upon detection, both sensor classes trigger innate inflammatory responses and allow the engagement of adaptive immunity^{1,2}. Endo-lysosomes are the entry sites for a variety of pathogens, and therefore the sites at which the immune system first senses their presence. Pathogens internalized by endocytosis are well known to activate TLRs 3 and 7–9 that are localized to endocytic compartments and detect ligands present in the endosomal lumen³. Internalized pathogens also activate sensors in the cytosol such as NOD1 and NOD2 (ref. 2), indicating that endosomes also provide for the translocation of bacterial components across the endosomal membrane. Despite the fact that NOD2 is well understood to have a key role in regulating innate immune responses and that mutations at the NOD2 locus are a common risk factor in inflammatory bowel disease and possibly other chronic inflammatory states^{4,5}, little is known about how its ligands escape from endosomes. Here we show that two endo-lysosomal peptide transporters, SLC15A3 and SLC15A4, are preferentially expressed by dendritic cells, especially after TLR stimulation. The transporters mediate the egress of bacterially derived components, such as the NOD2 cognate ligand muramyl dipeptide (MDP)^{6,7}, and are selectively required for NOD2 responses to endosomally derived MDP. Enhanced expression of the transporters also generates endosomal membrane tubules characteristic of dendritic cells, which further enhanced the NOD2-dependent response to MDP. Finally, sensing required the recruitment of NOD2 and its effector kinase RIPK2 (refs 8, 9) to the endosomal membrane, possibly by forming a complex with SLC15A3 or SLC15A4. Thus, dendritic cell endosomes are specialized platforms for both the luminal and cytosolic sensing of pathogens.

Although cytosolic NLRs monitor the presence of bacterial pathogens, only a minority of pathogenic or commensal bacteria reaches the cytosol, raising the question of how NLR ligands escape from endosomes after bacterial internalization. To determine if entry was required, we transfected 293 cells with NOD2 and an NF- κ B luciferase reporter, and cultured them with invasive *Salmonella typhimurium*, well known to activate the NOD2 pathway. As shown in Fig. 1a, wild-type *S. typhimurium* strongly induced NF- κ B even at low ratios of bacteria to 293 cells. Activation of NOD2 was dependent on host cell entry, because an invasion-defective *S. typhimurium* mutant (*AinvG*)¹⁰ did not induce NF- κ B (blue circles), similarly to a non-invasive bacterium (*Escherichia coli*; green circles). The failure of the *AinvG* mutant to trigger NOD2 signalling indeed reflected a failure of entry. First, blocking internalization of wild-type *S. typhimurium* by cytochalasin D also inhibited NOD2 activation (Fig. 1b), and without affecting signalling by soluble MDP added extracellularly (Extended Data Fig. 1a). Second, when the *AinvG* entry defect was complemented by expressing *Yersinia* invasin, which

triggers phagocytic entry by binding to β 1 integrins on target cells¹¹, NOD2 activation was completely restored (Extended Data Fig. 1b). After internalization, wild-type *S. typhimurium* was found within LAMP1-positive phagolysosomes (Extended Data Fig. 1c), indicating that ligands such as MDP do not gain access to NOD2 by phagosome disruption. Indeed, when the phagosome disruption occurs (as in the *sifA* mutant¹²), there is a marked activation of NOD2 even at very low bacterial exposures (Extended Data Fig. 1d–f). Although these observations indicate the involvement of a transporter under normal conditions, the relevant transporter was not bacterially encoded. NOD2 activation did not require the *S. typhimurium* effector proteins SipA and SopE (Extended Data Fig. 1g), recently suggested to activate NOD2 and MAP-kinase after translocation to the cytosol via the bacterial type III secretion system^{13,14}. More directly, phagocytosed MDP-coated 1- μ m latex beads were found to trigger NOD2-dependent activation of NF- κ B even in the absence of bacteria (Fig. 1c). Cytochalasin D as well as acidification inhibitors blocked NF- κ B activation and/or internalization of the beads (Fig. 1c, d and Extended Data Fig. 1h), strongly implying the involvement of endosomal transporters for MDP.

To characterize candidate transporters in endo-lysosomes, we performed quantitative SILAC (stable isotope labelling by amino acids in cell culture) proteomic analysis of endocytic compartments in dendritic cells. An enriched endo-lysosomal fraction was prepared by magnetic bead isolation after internalization of 200-nm iron oxide particles by isotopically labelled bone marrow-derived dendritic cells (BMDCs) \pm lipopolysaccharide (LPS) stimulation (Extended Data Fig. 2a–c). Out of 1,436 proteins identified in the endo-lysosome fractions, 8 and 45 were annotated as ATP-binding cassette (ABC) transporters and solute carrier family (SLC) transporters, respectively, in the UniProt database (Supplementary Table 1). Of these, five (ABCA1, ABCA3, SLC3A2, SLC7A2 and SLC15A3) were upregulated more than twofold by LPS treatment (Extended Data Fig. 2d), and showed a predominant localization to endocytic compartments (Extended Data Fig. 3a). The others were largely restricted to sites such as the endoplasmic reticulum (TAP1 and TAP2) or the plasma membrane (SLC7A11). Quantitative PCR revealed that only one candidate, SLC15A3, was expressed predominantly by dendritic cells. *SLC15A3* expression at the messenger RNA level increased 2.5-fold following LPS treatment of mouse BMDCs; it was also detected in a macrophage cell line (J774.2) (Fig. 1e, upper panel). In human cells, SLC15A3 was expressed by myeloid dendritic cells as well as by plasmacytoid dendritic cells (Fig. 1e, lower panel), but was not detected in other human or mouse cell types. SLC15A3 is a member of the proton-coupled peptide symporter family consisting of four genes, *SLC15A1* to *SLC15A4* (ref. 15). SLC15A4, a closely related family member, was also found in the endo-lysosome fraction albeit at lower induction by LPS treatment (Extended Data Fig. 2d). SLC15A3 and SLC15A4 were clearly localized to endosomes, lysosomes and endo-lysosomal tubules in BMDCs (Fig. 1g–i). Although SLC15A4 was not upregulated by LPS, it was constitutively highly expressed in plasmacytoid dendritic cells; it was also expressed at low levels in the 293 cells used for our reporter assays (Fig. 1f). The two other members of the family, SLC15A1 (also known as PepT1)

¹Genentech, 1 DNA Way, South San Francisco, California 94080, USA. ²Department of Cell Biology, University Medical Center Utrecht, 3584 CX Utrecht, The Netherlands.

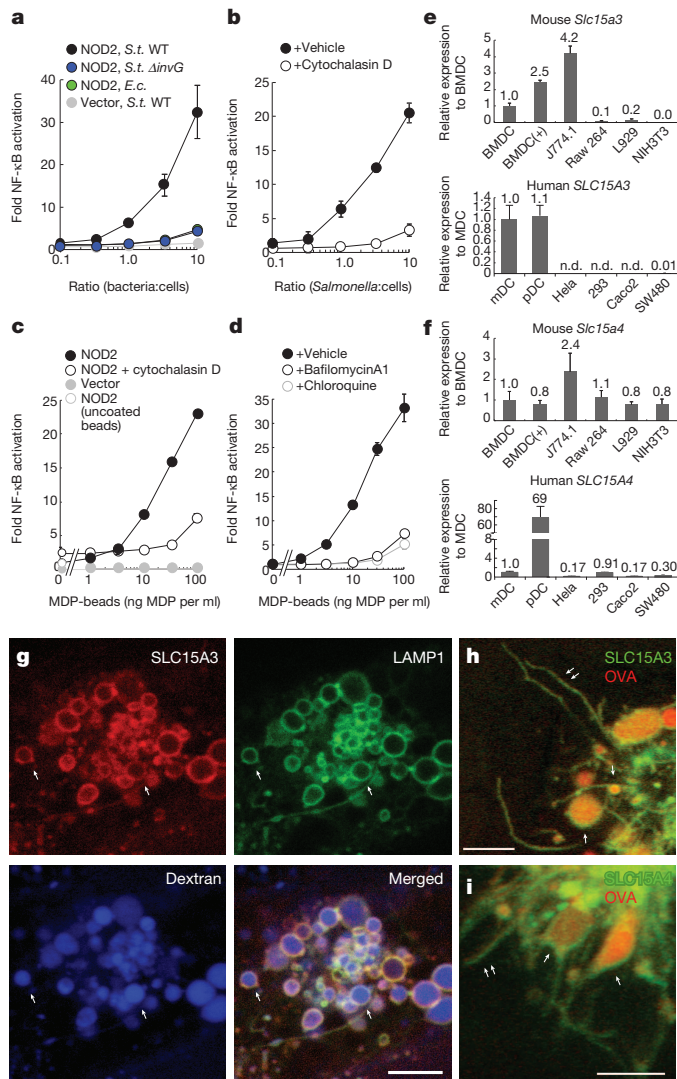


Figure 1 | Recognition of MDP from *Salmonella* and MDP-beads internalized into SLC15A3⁺ and SLC15A4⁺ endosomes. **a**, NF-κB-luciferase reporter assay of 293 cells transfected with an NF-κB-luciferase and either NOD2 or vector. Cells were cultured with *S. typhimurium* SL1344 (S.t. WT), *invG* mutant (*ΔinvG*), or *E. coli* BL21 (*E.c.*) for 1 h, then incubated overnight in fresh medium containing gentamycin. **b**, NOD2-transfected 293 cells were pre-incubated with or without cytochalasin D, then stimulated with *S. typhimurium*. **c**, NOD2- or vector-transfected cells were incubated with MDP-beads or uncoated beads (same amount to MDP-beads at 100 ng MDP ml⁻¹) with or without cytochalasin D. All data are presented as fold relative to unstimulated NOD2-transfected cells. Without NOD2 cDNA, NF-κB activation was negligible even in the presence of MDP. **d**, NOD2-transfected cells were pre-incubated with bafilomycin A1 or chloroquine. **e, f**, Expression of *SLC15A3* (**e**) and *SLC15A4* (**f**) in mouse and human cells was quantified by RT-PCR. (+) denotes LPS-stimulated dendritic cells. mDC, myeloid dendritic cells; pDC, plasmacytoid dendritic cells. **g–i**, Microscopy of BMDCs expressing SLC15A3-RFP and LAMP1-GFP (**g**), SLC15A3-GFP (**h**) or SLC15A4-GFP (**i**). Endo-lysosomes are labelled with dextran-Alexa647 or OVA-Alexa555. Single and double arrows mark vesicular and tubular endo-lysosomes. Bars, 5 μm. Unless noted otherwise data represent the mean of triplicates ± s.d. The data are representative of three independent experiments.

and SLC15A2 (also known as PepT2) localized to the plasma membrane, were not specific to dendritic cells, and were not found in the endo-lysosome fraction (Extended Data Fig. 3b, c and Supplementary Table 1).

SLC15A3 and SLC15A4 belong to a transporter family that co-transport di- or tri-peptides and protons across membrane using a proton gradient as the driving force¹⁵. Endosomes and lysosomes are acidic^{16,17}, suggesting that SLC15A3 and SLC15A4 transport small peptides to the cytosol

using the outward-directed proton gradient. We first asked if the expression of SLC15A3 enhanced the release of endocytosed MDP conjugated to rhodamine from endosomes. MDP-rhodamine was loaded into endosomes by endocytosis in cells expressing green-fluorescent-protein-conjugated SLC15A3-GFP or LAMP1-GFP (Fig. 2a). The rhodamine fluorescence in SLC15A3-GFP⁺ endosomes decreased more rapidly (Fig. 2b). We next expressed mutant transporters that localized to the plasma membrane following disruption of cytoplasmic di-leucine motifs¹⁸ required for endo-lysosomal localization (Fig. 2c and Extended Data Fig. 4a, b). When cells were incubated with MDP-rhodamine in acidic medium (pH 5.5), accumulation of MDP-rhodamine was significantly enhanced in cells expressing plasma membrane mutants of SLC15A3 (Fig. 2d). Only a slight increase in endocytosis was observed in cells expressing endosomal SLC15A3. Consistently, cells expressing the plasma membrane mutant were hyper-responsive to extracellular soluble MDP (Extended Data Fig. 4c, see also below). Thus, expression of transporter at cell surface increased uptake of free MDP from the culture medium, consistent with a directionality for transport of luminal MDP in endo-lysosomes to cytosol by wild-type transporters.

Transfection of a complementary DNA encoding wild-type SLC15A3 or SLC15A4 enhanced the NF-κB-luciferase signal induced by *S. typhimurium* or MDP-coated beads (Fig. 2e). Similar results were obtained using endogenous cytokine release (IL-6, TNF-α) as indicators of NOD2 signalling in Raw 264.7 macrophages, which express low levels of endogenous SLC15A3 and SLC15A4 (Fig. 2f and Extended Data Fig. 5a). SLC15A3 and SLC15A4 did not facilitate TLR and TNF-α signalling in 293 cells stimulated with flagellin and TNF-α (Extended Data Fig. 5b) or Raw 264.7 cells stimulated with LPS (Extended Data Fig. 5a), consistent with the specificity of the transporters for NOD2 ligands. Next, we generated BMDCs from a mouse strain deficient in SLC15A3 and assayed the ability of MDP-beads to elicit the cytokine release. When wild-type dendritic cells were incubated with MDP-beads, IL-6 was produced in a dose-dependent manner (Fig. 2g, closed bars). However, the production of IL-6 by SLC15A3-deficient dendritic cells was substantially lower than wild-type cells (open bars). Similarly, the production of IL-1β was also substantially lower in SLC15A3-deficient dendritic cells (Extended Data Fig. 5c). These results indicate that SLC15A3 facilitates the recognition of internalized MDP. SLC15A3-deficient BMDCs took up similar numbers of MDP-beads compared to wild-type, indicating that endocytic uptake was unaffected by the absence of SLC15A3 (Extended Data Fig. 5d). LPS- or *Salmonella*-induced production of IL-6 was also unaffected in SLC15A3-deficient dendritic cells (Extended Data Fig. 5e). The residual signalling seen in the knockout dendritic cells probably reflected the residual expression of SLC15A4 in these cells. The SLC15A3-deficient mice themselves exhibited no overt phenotype, at least under steady state conditions.

One of the most striking features of dendritic cell endo-lysosomes is their ability to extend long tubules following LPS treatment^{19–21}. Indeed, endo-lysosomes isolated from LPS-stimulated BMDCs contained numerous tubular structures (Extended Data Fig. 2b, arrows). Because SLC15A3 was strongly upregulated upon LPS stimulation, we asked if these two events were related. Remarkably, cells transfected with SLC15A3 exhibited abundant tubular structures compared to untransfected cells (Fig. 3a). Expression of SLC15A4 also generated tubules (Fig. 3b), although other endo-lysosomal transporters did not (see Extended Data Fig. 3a). Tubule induction was efficient, with 53% and 47% of 293 cells expressing SLC15A3 and SLC15A4 exhibiting tubules (Fig. 3c). Conversely, in BMDCs from *Slc15a3*^{-/-} mice, fewer than 30% of total cells exhibited tubules after LPS treatment, whereas ~80% of wild-type BMDCs exhibited tubular endo-lysosomes (Fig. 3e). The reduced tubulation was rescued by expressing SLC15A3-GFP, but not by the GFP empty vector (Fig. 3d). Taken together, these gain and loss-of-function experiments provide strong evidence that the induction of SLC15A3 and SLC15A4 in response to stimuli such as LPS results in the formation of endo-lysosomal tubules. We asked if tubulation itself might have significance for NOD2 sensing. Because the extensions tracked along cytoplasmic microtubules

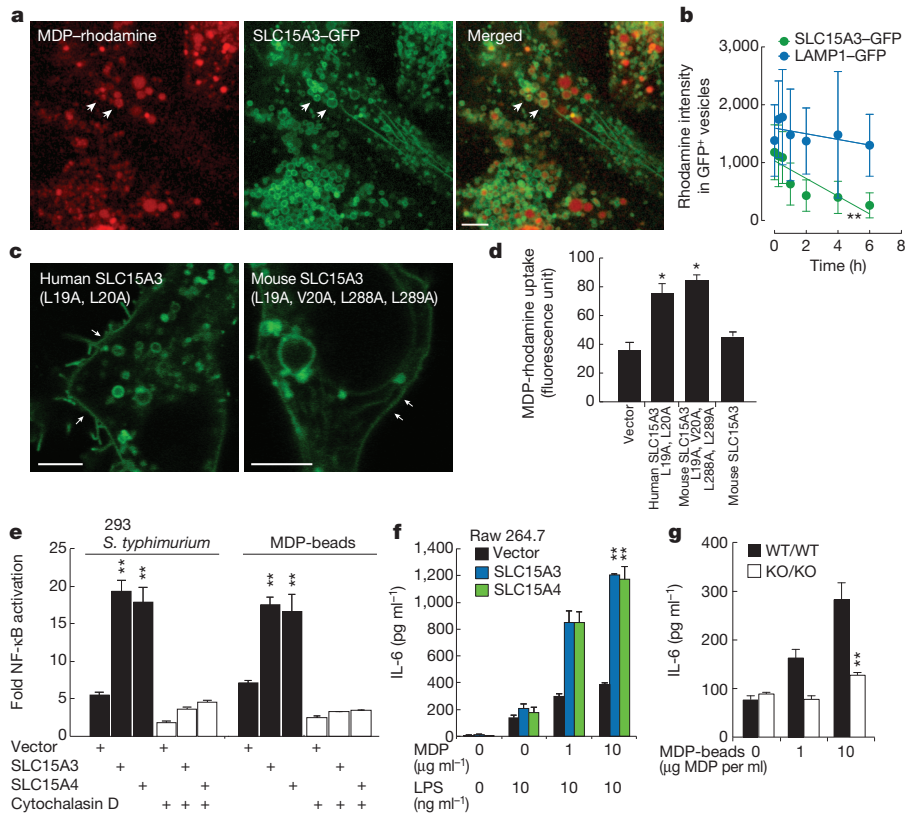
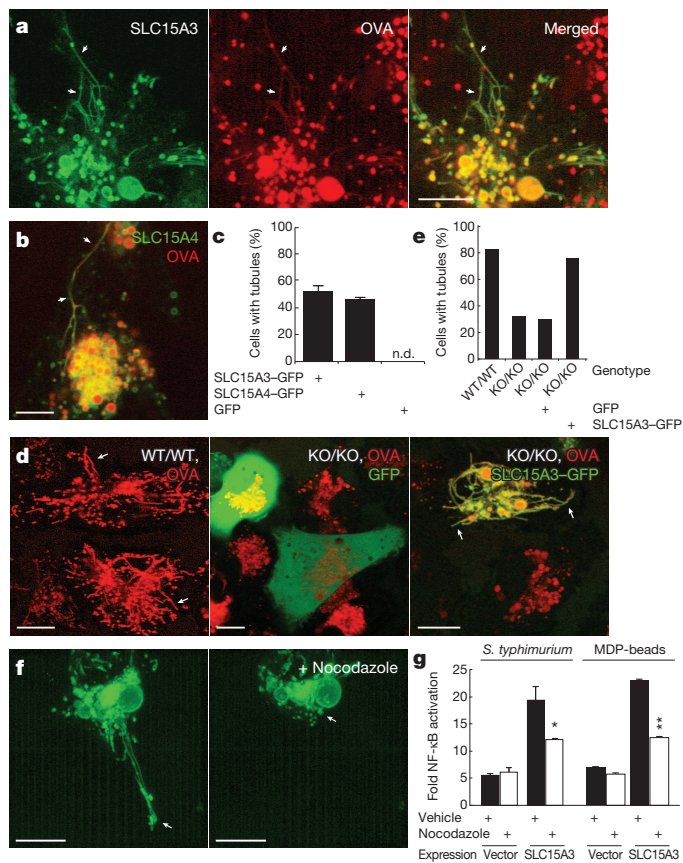


Figure 2 | SLC15A3 and SLC15A4 facilitate the recognition of phagosomal MDP. **a**, **b**, 293 cells expressing SLC15A3-GFP (**a**) or LAMP1-GFP (**b**) were incubated with MDP-rhodamine, then chased in fresh medium. Rhodamine in GFP-positive endosomes (arrows) were quantified by fluorescence microscope and ImageJ (**b**). Means \pm s.d. **Two regression slopes differ significantly ($P = 0.006$, $n = 20$ endosomes; t -test). Bars, 5 μm . **c**, Microscopy of di-leucine mutants of human (left) or mouse (right) SLC15A3-GFP in 293 cells. Arrows indicate cell surface. **d**, Cells transfected with SLC15A3, di-leucine mutants or vector were incubated with soluble MDP-rhodamine in medium (pH 5.5) for 30 min, then the uptake was quantified by a fluorophotometre. **e**, 293 cells transfected with an NF- κ B-luciferase, NOD2, and either SLC15A3, SLC15A4 or vector, are incubated with *S. typhimurium* (multiplicity of infection = 10) or MDP-beads ($1 \mu\text{g MDP ml}^{-1}$). White bars represent results for the same conditions but with the addition of cytochalasin D. **f**, Raw 264.7 cells stably expressing SLC15A3 or SLC15A4 were stimulated with MDP plus 10 ng ml^{-1} LPS for 24 h. **g**, SLC15A3-deficient or wild-type BMDCs were incubated with MDP-beads. IL-6 in medium was quantified by ELISA. Means \pm s.d. * $P < 0.05$ and ** $P < 0.01$ by t -test.

(Extended Data Fig. 6), it was not surprising that nocodazole treatment, which inhibits the polymerization of microtubules, collapsed the tubular network (Fig. 3f). Interestingly, nocodazole also reduced the activation



of NF- κ B enhanced by SLC15A3 in cells simulated with *Salmonella* or MDP-beads (Fig. 3g, SLC15A3), although this enhancement did not occur in cells where tubules were not present, (Fig. 3g, vector). Thus, endo-lysosomal tubules generated by transporter expression further enhance MDP escape or detection, perhaps by increasing the surface area of the compartment^{22,23}.

We next determined that there was a spatial relationship between NOD2 and bacteria-containing phagosomes. Because endogenous NOD2 is expressed at low levels, 293 cells expressing red-fluorescent-protein-conjugated RFP-NOD2 were used for these experiments, the fusion protein having been found functional for NF- κ B signalling (see Extended Data Fig. 7e). NOD2 was associated with *Salmonella*-containing phagosomes (Fig. 4a, b), whose membranes were also positive for LAMP1 and SLC15A4 (Fig. 4c and Extended Data Fig. 7a). Additional pools of NOD2 were found associated with the plasma membrane and diffusely distributed in the cytosol (Fig. 4a), consistent with earlier work that suggested NOD1 recruitment at sites of *Shigella flexneri* attachment²⁴. NOD1, which recognizes γ -D-glutamyl-diaminopimelic acid in peptidoglycan^{25,26}, was also found on *Salmonella*-containing phagosomes (Extended Data Fig. 7b). Association of NOD2 with the plasma membrane requires a

Figure 3 | SLC15A3 and SLC15A4 generate endo-lysosomal tubules. **a**, **b**, Microscopy of 293 cells expressing SLC15A3-GFP or SLC15A4-GFP. Arrows mark tubules with OVA-Alexa555. **c**, Frequencies of cells with endo-lysosomal tubules in the transfected 293 cells. Means \pm s.d. ($n = 80$ cells in triplicates). n.d., not detected. **d**, Microscopy of wild-type or SLC15A3-deficient BMDCs stimulated with LPS. GFP or SLC15A3-GFP was expressed in SLC15A3 knockout BMDCs. **e**, Frequencies of BMDCs with tubules. For SLC15A3-deficient cells, percentages were calculated as a function of GFP-positive cells. **f**, 293 cell expressing SLC15A3-GFP was imaged before and after 10 μM nocodazole treatment (17 min). Arrows indicate tubules. **g**, 293 cells transfected with NOD2 and either SLC15A3 or vector, were pre-incubated with or without nocodazole, then stimulated with *S. typhimurium* or MDP-beads. Means \pm s.d. * $P < 0.05$ and ** $P < 0.01$. Bars, 10 μm (**a**, **b**, **d**) and 5 μm (**f**).

leucine-rich repeat domain found at the NOD2 carboxy terminus²⁷. A frameshift mutation (3020insC) that causes a truncation of the leucine-rich domain is a risk factor for inflammatory bowel disease^{5,28}, suggesting that the membrane association of NOD2 is important for pathogen sensing. Indeed, the 3020insC variant was not associated with *Salmonella*-containing phagosomes, but was largely found in the cytosol (Fig. 4d). These results were confirmed by immunoelectron microscopy (Fig. 4b). In cells expressing RFP–NOD2, 0.88 gold particles per phagosome were detected whereas in cells expressing mutant NOD2–3020insC, only 0.22 gold particles per phagosome were found (*t*-test, *P* = 0.001); the residual association in mutant-expressing cells probably reflected the physical proximity of unbound soluble NOD2–3020insC. Cell fractionation showed that full-length NOD2 co-fractionated with the late endosomal-lysosomal marker LAMP1 or LAMP2: ~20% of RFP– or Flag–NOD2 was found in the phagosome fraction of 293 and BMDs, compared with only 4% of tagged NOD2–3020insC (Extended Data Fig. 7c). Consistently, NOD2 was associated with the LAMP2-positive phagosomes that contain internalized MDP-beads in BMDs (Extended Data Fig. 7d). Importantly, cells expressing mutant NOD2–3020insC were largely deficient at NF- κ B activation in response to *S. typhimurium* entry (Extended

Data Fig. 7e). Thus, the association of NOD2 with pathogen-containing endocytic compartments is an important feature of MDP sensing.

We next asked if NOD2 might be in a complex with SLC15A3 or SLC15A4. Indeed, NOD2 was detected in immunoprecipitates of both SLC15A3 and SLC15A4 (Fig. 4e, left); similarly, SLC15A3 and SLC15A4 could also be detected in NOD2 immunoprecipitates. Importantly, the abundant endo-lysosomal protein LAMP1 was not associated with NOD2, providing support for the specificity of the SLC15–NOD2 association (Fig. 4e, right). Consistent with localization of NOD2–3020insC variant in the cytosol (Fig. 4d), NOD2–3020insC protein was greatly reduced in the immunoprecipitates of SLC15A3 and SLC15A4 (Fig. 4e, left). SLC15A3 and SLC15A4 co-immunoprecipitated SLC15A4 and SLC15A3, respectively (Fig. 4f), indicating that SLC15A3 and SLC15A4 can form heterodimers with each other. Finally, NOD2 binds an effector protein kinase RIPK2, which is essential for activating downstream signalling via NF- κ B and microRNA-29^{8,9,29}. When RFP–NOD2 was expressed with GFP–RIPK2 in 293 cells, the RIPK2 was found in endo-lysosomes that contained NOD2 (Fig. 4g). In the absence of NOD2 expression, GFP–RIPK2 was low or absent in endosomes (Extended Data Fig. 7f), indicating that RIPK2 is recruited to endosomes by NOD2, suggesting

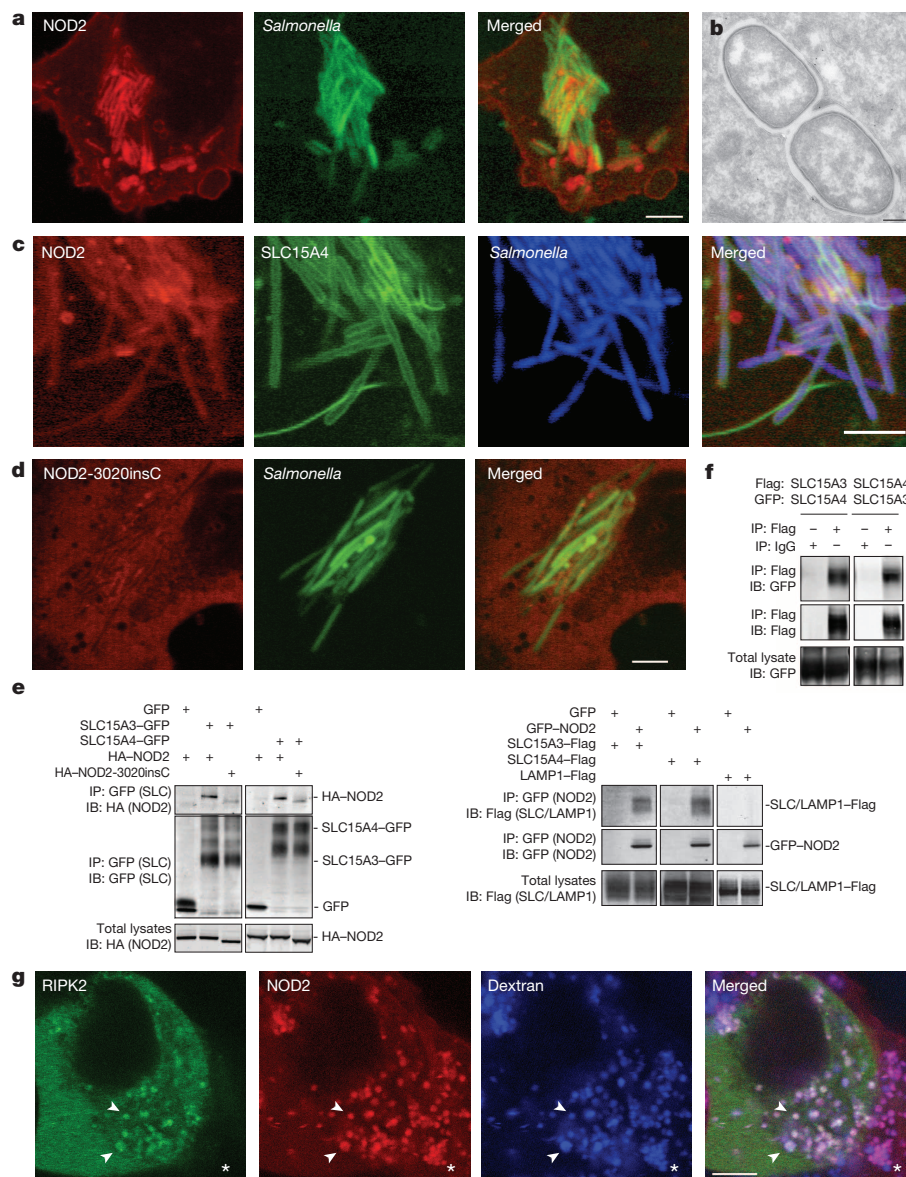


Figure 4 | NOD2 forms a sensor complex on endosomes. **a–d**, Fluorescence microscopy (**a**, **c**, **d**) or immuno-electron microscopy (**b**) of NOD2 in *Salmonella*-infected 293 cells. Cells stably expressing RFP–NOD2 (**a–c**) or RFP–NOD2–3020insC (**d**) and SLC15A4–GFP (**c**) were incubated with *S. typhimurium* expressing GFP (**a**, **b**, **d**) or tagBFP (**c**) as in Fig. 1. RFP–NOD2 in **b** was identified by anti-RFP antibody and protein A–gold (10 nm). **e**, **f**, Co-immunoprecipitation of NOD2, SLC15A3 and SLC15A4. Tagged proteins were expressed in 293 cells, isolated by anti-GFP (**e**) or anti-Flag (**f**) beads (immunoprecipitation, IP), then immunoblotted (IB). **g**, Fluorescence microscopy of 293 cells expressing either GFP–RIPK2 or RFP–NOD2. Endo-lysosomes are labelled with dextran-Alexa647 (arrowheads). Asterisk (*) indicates a cell expressing RFP–NOD2 but not GFP–RIPK2. Bars, 5 μ m (**a**, **c**, **d**, **g**) and 0.2 μ m (**b**).

that membrane transporters, NOD2 and the effector protein kinase contribute to the assembly of the signalling complex on endosomes.

Our genetic and functional analysis has demonstrated that MDP transport from endosomes and lysosomes was dependent on two dendritic-cell- and macrophage-specific endo-lysosomal transporters, SLC15A3 and SLC15A4. Importantly, MDP sensing also required the recruitment of NOD1 and NOD2, and accordingly RIPK2, to transporter-containing compartments. Thus, endo-lysosomal membranes in dendritic cells (and macrophages) are not only specialized to enable the selective escape of key bacterial components, but also to serve as platforms that assemble critical microbial sensors and downstream effectors. Together with the fact that the transporters induce a marked tubulation of endosomal compartments to enhance the sensitivity of MDP detection, our results establish the conceptual importance of endosomes as signalling platforms specialized for triggering innate immune responses.

METHODS SUMMARY

MDP-beads were prepared by mixing tosyl-activated Dynal MyOne beads (Invitrogen) and MDP-rhodamine B (InvivoGen). For reporter assay, 293 cells (40×10^3) stably transfected with pNIFTY2-Luc and pUNO-hNOD2a (InvivoGen) were used. For co-expression with SLC15 transporters, 293 cells were transiently transfected with pNIFTY2-Luc, pUNO-hNOD2a, and either plasmid of EGFP, SLC15A3-EGFP, or SLC15A4-EGFP using Lipofectamine 2000. Cells were incubated with MDP or MDP-beads for 7 h, or *S. typhimurium* (SL1344) or *E. coli* (BL21) for 1 h. The cells incubated with bacteria were washed with fresh medium, then cultured in the presence of $20 \mu\text{g ml}^{-1}$ gentamycin for 24 h. Luciferase activity was expressed as fold activation relative to the activity of unstimulated NOD2-transfected cells. BMDCs were differentiated from bone marrow using GM-CSF. Cytokine secretion from BMDCs (200×10^3) or Raw 264.7 (100×10^3) were measured by ELISA. The amount of beads in each culture was adjusted by adding non-coated beads.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 20 January 2013; accepted 6 February 2014.

Published online 30 March; corrected online 7 May 2014 (see full-text HTML version for details).

1. Kimbrell, D. A. & Beutler, B. The evolution and genetics of innate immunity. *Nature Rev. Genet.* **2**, 256–267 (2001).
2. Elinav, E., Strowig, T., Henao-Mejia, J. & Flavell, R. A. Regulation of the antimicrobial response by NLR proteins. *Immunity* **34**, 665–679 (2011).
3. Kawai, T. & Akira, S. Toll-like receptors and their crosstalk with other innate receptors in infection and immunity. *Immunity* **34**, 637–650 (2011).
4. Hugot, J. P. *et al.* Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn's disease. *Nature* **411**, 599–603 (2001).
5. Ogura, Y. *et al.* A frameshift mutation in NOD2 associated with susceptibility to Crohn's disease. *Nature* **411**, 603–606 (2001).
6. Girardin, S. E. *et al.* Nod2 is a general sensor of peptidoglycan through muramyl dipeptide (MDP) detection. *J. Biol. Chem.* **278**, 8869–8872 (2003).
7. Inohara, N. *et al.* Host recognition of bacterial muramyl dipeptide mediated through NOD2. Implications for Crohn's disease. *J. Biol. Chem.* **278**, 5509–5512 (2003).
8. Ogura, Y. *et al.* Nod2, a Nod1/Apaf-1 family member that is restricted to monocytes and activates NF- κ B. *J. Biol. Chem.* **276**, 4812–4818 (2001).
9. Kobayashi, K. *et al.* RICK/Rip2/CARDIAK mediates signalling for receptors of the innate and adaptive immune systems. *Nature* **416**, 194–199 (2002).

10. Collazo, C. M. & Galan, J. E. The invasion-associated type-III protein secretion system in *Salmonella*—a review. *Gene* **192**, 51–59 (1997).
11. Isberg, R. R., Voorhis, D. L. & Falkow, S. Identification of invasins: a protein that allows enteric bacteria to penetrate cultured mammalian cells. *Cell* **50**, 769–778 (1987).
12. Beuzón, C. R. *et al.* *Salmonella* maintains the integrity of its intracellular vacuole through the action of SifA. *EMBO J.* **19**, 3235–3249 (2000).
13. Keestra, A. M. *et al.* A *Salmonella* virulence factor activates the NOD1/NOD2 signaling pathway. *MBio* **2**, e00266–11 (2011).
14. Hardt, W. D., Chen, L. M., Schuebel, K. E., Bustelo, X. R. & Galan, J. E. *S. typhimurium* encodes an activator of Rho GTPases that induces membrane ruffling and nuclear responses in host cells. *Cell* **93**, 815–826 (1998).
15. Daniel, H. & Kottra, G. The proton oligopeptide cotransporter family SLC15 in physiology and pharmacology. *Eur. J. Phys.* **447**, 610–618 (2004).
16. Mellman, I., Fuchs, R. & Helenius, A. Acidification of the endocytic and exocytic pathways. *Annu. Rev. Biochem.* **55**, 663–700 (1986).
17. Nakamura, N., Tanaka, S., Teko, Y., Mitsui, K. & Kanazawa, H. Four Na⁺/H⁺ exchanger isoforms are distributed to Golgi and post-Golgi compartments and are involved in organelle pH regulation. *J. Biol. Chem.* **280**, 1561–1572 (2005).
18. Bonifacio, J. S. & Traub, L. M. Signals for sorting of transmembrane proteins to endosomes and lysosomes. *Annu. Rev. Biochem.* **72**, 395–447 (2003).
19. Boes, M. *et al.* T-cell engagement of dendritic cells rapidly rearranges MHC class II transport. *Nature* **418**, 983–988 (2002).
20. Chow, A., Toomre, D., Garrett, W. & Mellman, I. Dendritic cell maturation triggers retrograde MHC class II transport from lysosomes to the plasma membrane. *Nature* **418**, 988–994 (2002).
21. Kleijmeer, M. *et al.* Reorganization of multivesicular bodies regulates MHC class II antigen presentation by dendritic cells. *J. Cell Biol.* **155**, 53–64 (2001).
22. Chow, A. Y. & Mellman, I. Old lysosomes, new tricks: MHC II dynamics in DCs. *Trends Immunol.* **26**, 72–78 (2005).
23. Marsh, M., Griffiths, G., Dean, G. E., Mellman, I. & Helenius, A. Three-dimensional structure of endosomes in BHK-21 cells. *Proc. Natl Acad. Sci. USA* **83**, 2899–2903 (1986).
24. Kufer, T. A., Kremmer, E., Adam, A. C., Philpott, D. J. & Sansonetti, P. J. The pattern-recognition molecule Nod1 is localized at the plasma membrane at sites of bacterial interaction. *Cell. Microbiol.* **10**, 477–486 (2008).
25. Chamillard, M. *et al.* An essential role for NOD1 in host recognition of bacterial peptidoglycan containing diaminopimelic acid. *Nature Immunol.* **4**, 702–707 (2003).
26. Girardin, S. E. *et al.* Nod1 detects a unique muropeptide from gram-negative bacterial peptidoglycan. *Science* **300**, 1584–1587 (2003).
27. Barnich, N., Aguirre, J. E., Reinecker, H. C., Xavier, R. & Podolsky, D. K. Membrane recruitment of NOD2 in intestinal epithelial cells is essential for nuclear factor- κ B activation in muramyl dipeptide recognition. *J. Cell Biol.* **170**, 21–26 (2005).
28. Rivas, M. A. *et al.* Deep resequencing of GWAS loci identifies independent rare variants associated with inflammatory bowel disease. *Nature Genet.* **43**, 1066–1073 (2011).
29. Brain, O. *et al.* The intracellular sensor NOD2 induces microRNA-29 expression in human dendritic cells to limit IL-23 release. *Immunity* **39**, 521–536 (2013).

Supplementary Information is available in the online version of the paper.

Acknowledgements The authors thank Mellman laboratory members for discussions, C. Chalouni for excellent advice and assistance with fluorescence microscopy, S. van Dijk and R. Scriwanek for immuno-electron microscopy, and R. Gentlemen, V. Dixit and N. Kayagaki for advice. N.N., J.R.L., Q.P., Z.J., C.B., M.S., L.D. and I.M. are full-time employees of Genentech, which funded the research in its entirety.

Author Contributions N.N., L.D. and I.M. designed the experiments. N.N. performed the experiments. J.R.L. and Q.P. ran mass spectrometry. N.N., Z.J. and C.B. analysed gene expression and mass spectrometry data. A.D.M. and J.K. performed electron microscopy. M.S. generated mouse lines. N.N. and I.M. wrote the manuscript.

Author Information Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to I.M. (mellman.ira@gene.com).

METHODS

Materials. MDP and *E. coli* LPS were obtained from InvivoGen and Sigma. MDP-beads were prepared by mixing tosyl-activated Dynal MyOne beads (Invitrogen) and MDP-rhodamine B (InvivoGen). The beads absorbed 4.4 fg MDP-rhodamine per bead. Antibodies were purchased from following companies: LAMP1 (25) and syntaxin-6 (30), BD Bioscience; Na⁺/K⁺-ATPase α 1 subunit (C464.6), Upstate; transferrin receptor (H68.4), Zymed; GAPDH (6C5), Ambion; Flag-tag (M2), Sigma. Polyclonal antibodies against GFP (632459) and tagRFP (AB233) were obtained from Clontech and Evrogen. Anti-MHCII IAB polyclonal and anti-LAMP-2 monoclonal (clone GL2A7) antibodies are described previously³⁰.

Animals and cell culture. Research involving animals complied with protocols approved by the Genentech Institutional Animal Care and Use Committee. BMDCs were differentiated from bone marrow of 7–12 weeks old mice using GM-CSF (peprotech)³¹. BMDCs expressing MHCII-GFP were prepared from MHCII-GFP knock-in mice¹⁹. Human dendritic cells were obtained from blood³². All procedures used in this study complied with federal guidelines and institutional policies of the Genentech Institutional Animal Care and Use Committee. Gene expression was analysed using TaqMan RT-PCR: SLC15A1 (assay ID: Hs00192639_m1), SLC15A2 (assay ID: Hs00221539_m1), SLC15A3 (assay IDs: Hs00275455_m1 and Mm00491666_m1), SLC15A4 (assay IDs: Hs00377326_m1 and Mm00505709_m1), and GAPDH (assay IDs: Hs99999905_m1 and Mm99999915_g1) were obtained from Applied Biosystems. *S. typhimurium* and *E. coli* were cultured in LB at 37 °C. Overnight culture of bacteria was diluted 20-fold in LB containing 0.3 M NaCl, and incubated for 2 h before use. *InvG*, *SifA*, *SipA* and *SopE* genes were disrupted using a λ -Red plasmid pSIM18³³. Targeting cassettes that comprise kanamycin-resistance gene flanked by LoxP sites and homology arms, were generated by PCR using pBlight-TK³⁴ and the following primers: *InvG*: targeting vector, 5'-agtggatcaaaagagctgacggctgttcgcgcggaaattcaaatattattcaattggcagacaatgaataactctgataatgtatgtctatcgaagtattctgttcaaatatcatcgccat-3' and 5'-gatgagatgtctgcgggacaatattctggaaatgaataacccggaggttgagccaggaatcaattatgataactctgataatcgaagtattcagaagaactctgcaagaagcgc-3'; genotyping, 5'-caccataatgacgcttaacatcgg-3' and 5'-ggcctgtgttattccgctgttg-3'. *SifA*: targeting vector, 5'-cgcatgttgagataaaaagggctcgatttaataatgtatgtcattttaccagataatgtagatataaactctgataatgtatgtctatcgaagtattctgttgacaataatcatcgccat-3' and 5'-actcagtagtgcacaaagagtgagtagaacctgaacgtgacgtctgagaagcgtcgtctgatttataatctcgtatagcagataattacgaagtattcagaagaactctgcaagaagcgc-3'; genotyping, 5'-tgctgttaacgcaagctctcag-3' and 5'-gactattctcatcgatccggtca-3'. *SopE*: targeting vector, 5'-tgctataaaaataaaagcattctcgtatctatataaataaagtattgtatcatataaaagatcattaccataactctgataatgtatgatacgaagtattctgttgacaataatcatcgccat-3' and 5'-ggggaaattatggataacttggttcatataatcagaagggctcgcgataatttttgggttttcagtgataactctgatacagattatcagaagtattcagaagaactctgcaagaagcgc-3'; genotyping, 5'-attgaatcagcgaatttttcaat-3' and 5'-gacggttttagctccgagtagttgc-3'. *SipA*: targeting vector, 5'-agcagtagcataagtagcagcctggaaacccgaagcttctcgaagataacagaagaggaatataataaactctgataatgtatgatacgaagtattcgcggaacccctattgttt-3' and 5'-aatacagaaggtgacatcttttgaactctgcttcaatattcattcagatctttccggttaataaactctgatacagattatgtagtctgacagattaccatgc-3'; genotyping, 5'-aaaaatagcaatgcaattcattgtacg-3' and 5'-agatctcaaccagatgggtctgac-3'. The kanamycin-resistance gene was removed from the chromosome by transforming 705-Cre (Gene Bridges). Gene disruption was confirmed by genomic PCR. A low copy plasmid, pSTV29 (Takara) was used to express EGFP, tagBFP, *SifA* and *invasin* gene of *Y. pseudotuberculosis* (NCTC 10275). *SifA* and *invasin* genes were amplified from genomic DNAs of *S. typhimurium* and *Y. pseudotuberculosis*, respectively, by PCR using primers: *SifA*, 5'-gggagatctacacgcatcagcgaattttatc-3' and 5'-ggggaattctataaaaacaataaacagccgctttg-3'; *Invasin*, 5'-ggggagatccgtatcatgaacggataccagac-3' and 5'-ggcggatcaacgctgggtattgc-3'. Bacteria were transformed using Gene Pulser (Bio-Rad).

Reporter assay. 293 cells (40×10^3) stably transfected with pNIFTY2-Luc and pUNO-hNOD2a (InvivoGen) were used for the reporter assay. For co-expression with SLC15 transporters, 293 cells were transiently transfected with pNIFTY2-Luc (25 ng), pUNO-hNOD2a (1 ng), and either plasmid of EGFP, SLC15A3-EGFP, or SLC15A4-EGFP (1.8 ng) using Lipofectamine 2000. Cells were incubated with MDP or MDP-beads for 7 h, or *S. typhimurium* or *E. coli* for 1 h. Cells incubated with bacteria were washed with fresh medium, then cultured in the presence of 20 μ g ml⁻¹ gentamycin for 24 h. Cytochalasin D (10 μ M, 1 h), nocodazole (10 μ M, 30 min), bafilomycin A1 (1 μ M, 30 min) or chloroquine (10 μ M, 30 min) was added before the stimulation. Luciferase activity was measured by Dual-Glo luciferase assay kit (Promega), then expressed as fold activation relative to the activity of unstimulated NOD2-transfected cells. Production of IL-6, IL-1 β , and TNF- α from BMDCs (200×10^3) or Raw 264.7 (100×10^3) were measured by ELISA (R&D systems). The amount of beads in each culture was adjusted by adding non-coated beads.

Mass spectrometry of endo-lysosomes. SILAC was performed as described previously^{35–37}. BMDCs were cultured in heavy lysine-¹³C₆, ¹⁵N₂ and arginine-¹³C₆, ¹⁵N₄ or light lysine-¹²C₆, ¹⁴N₂ and arginine-¹²C₆, ¹⁴N₂ for 7 days. Cells were incubated

with 0.3 mg ml⁻¹ iron oxide particles of 200-nm diameter (Ademtech) and ovalbumin-Alexa555 for 2.5 h, then stimulated with (heavy labelled) or without (light labelled) LPS (30 ng ml⁻¹) for 3 h. Cells were lysed in 10 mM HEPES-KOH, pH 7.4, containing 250 mM sucrose and, 1 mM EDTA (Total), then spun at 120g. The post-nuclear supernatants (PNS) were applied to a MACS MS column on a magnet holder (Miltenyi). Endo-lysosomes were eluted from the column by separating the magnet. Heavy and light labelled endosomal proteins were mixed in a 1:1 ratio, resolved by SDS-PAGE, then treated with trypsin. Peptides were separated on a NanoAcquity UPLC (Waters), and analysed using an LTQ XL-Orbitrap mass spectrometer (ThermoFisher). Precursor ions were analysed in the FTMS at a resolution of 60,000 $M/\Delta M$ at m/z 400. Tandem mass spectral data was submitted for Mascot database (Matrix Science)³⁸, then quantified using VistaQuant as previously described³⁹.

Fluorescence microscopy. Microscopic images were taken using live cells unless specified. GFP- and RFP-fusion proteins were expressed in BMDC using retrovirus produced by pBMN-Z or pQCXIP and Phoenix-Eco or GP2-293. To label endo-lysosomes, cells were incubated with 20 μ g ml⁻¹ of ovalbumin- or dextran-Alexa conjugate (Invitrogen), then chased with fresh medium for 30 min. BMDCs were stimulated with LPS (30 ng ml⁻¹) before microscopy. To measure MDP-rhodamine in endo-lysosomes, 293 cells expressing SLC15A3-GFP or LAMP1-GFP were incubated with MDP-rhodamine (1 μ g ml⁻¹) overnight, then chased with fresh medium without MDP-rhodamine on Leica SP5. MDP-rhodamine in GFP-positive endosomes were quantified by ImageJ.

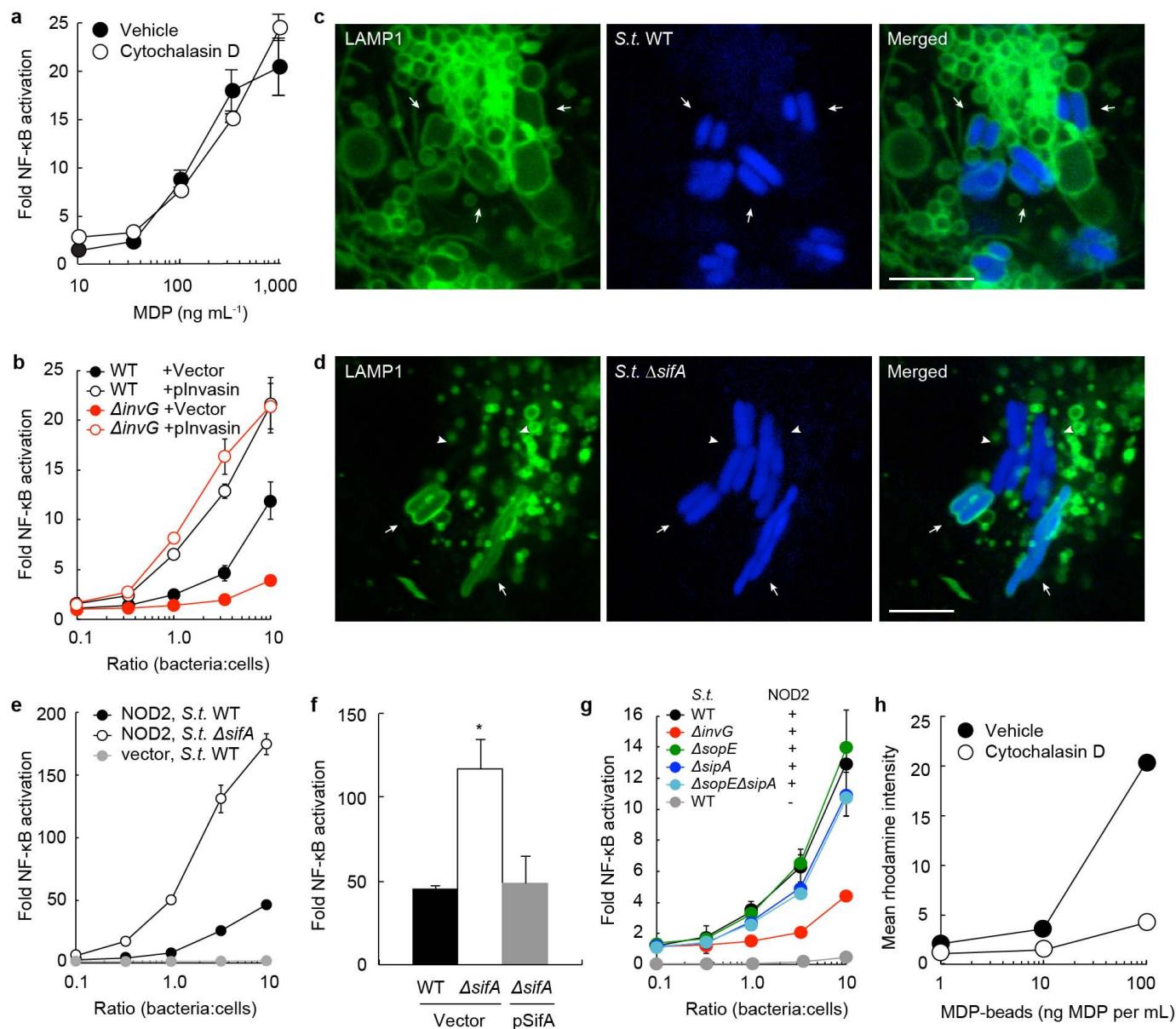
MDP-uptake assay. COS7 cells (0.16×10^6) were transfected with expression plasmid of either GFP, SLC15A3-GFP or the di-leucine mutants of SLC15A3-GFP, then cultured overnight. Cells were washed with transport buffer (5 mM D-glucose, 140 mM NaCl, 1 mM MgSO₄, 20 mM potassium phosphate, pH 7.4), then incubated with 10 μ g ml⁻¹ MDP-rhodamine in the buffer at pH 5.5 at 37 °C for 30 min. Cells were washed with ice-cold buffer, then lysed in 0.1% Triton X-100 and 0.1% NP-40. MDP-rhodamine was quantified by a fluorophotometer.

Electron microscopy. The isolated endo-lysosome fraction was fixed in 2% glutaraldehyde, post-fixed in 2% OsO₄ and en bloc stained in 1% uranyl acetate. For immunoelectron microscopy, cells were fixed with 4% paraformaldehyde, embedded in 12% gelatin, cryoprotected with 2.3 M sucrose, then frozen in liquid nitrogen. Ultrathin cryosections on copper grids were incubated with anti-tagRFP antibody followed by protein-A gold, as described earlier⁴⁰. The sections were then contrasted with a 1.8% methylcellulose, 0.4% uranyl acetate mixture.

Immunoprecipitation. 293 cells (3×10^6) were transfected with empty or expression plasmids of NOD2, SLC15A3, SLC15A4, or LAMP1, lysed in a buffer containing 40 mM TrisCl (pH 7.4), 0.5% NP-40, 50 mM NaCl, and 5 mM MgCl₂, then applied to μ MACS GFP Isolation Kit (Miltenyi) or EZview anti-Flag M2 affinity gel (Sigma) according to the manufacturer's instruction.

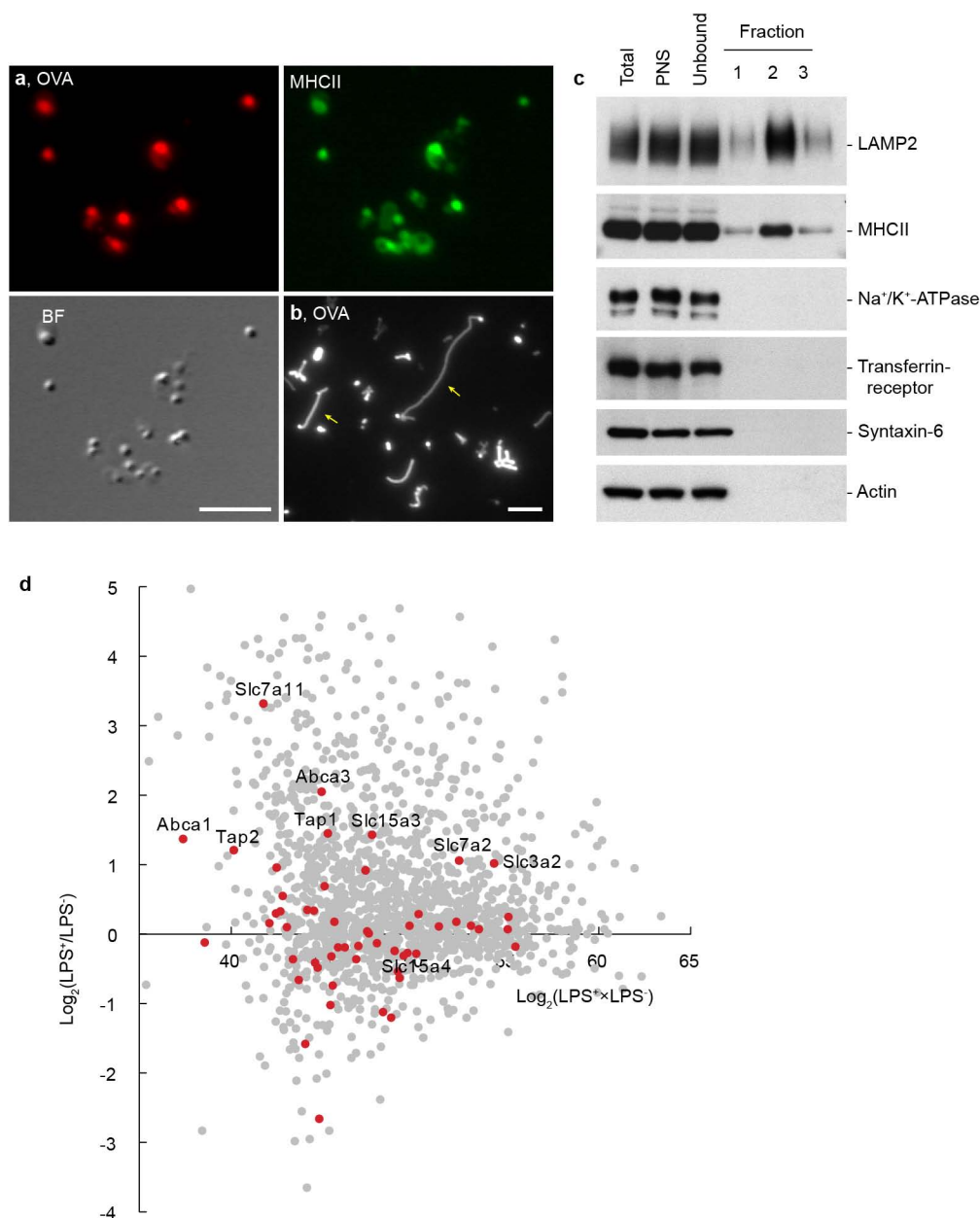
SLC15A3 knock-out mouse. Mice carrying a floxed *Slc15a3* were generated from an embryonic stem cell clone (*Slc15a3^{tm1a(KOMP)Wtsi}*) of Knockout Mouse Project Repository, then intercrossed with C57BL/6-Gt(ROSA)26Sor^{tm16(Cre)Arte} Cre strain (Taconic) to produce knockout allele. Homozygous knockout mice are born at the Mendelian frequencies.

- Ma, J. K., Platt, M. Y., Eastham-Anderson, J., Shin, J. S. & Mellman, I. MHC class II distribution in dendritic cells and B cells is determined by ubiquitin chain length. *Proc. Natl. Acad. Sci. USA* **109**, 8820–8827 (2012).
- Shin, J. S. *et al.* Surface expression of MHC class II in dendritic cells is controlled by regulated ubiquitination. *Nature* **444**, 115–118 (2006).
- Smed-Sørensen, A. *et al.* Differential susceptibility to human immunodeficiency virus type 1 infection of myeloid and plasmacytoid dendritic cells. *J. Virol.* **79**, 8861–8869 (2005).
- Chan, W. *et al.* A recombineering based approach for high-throughput conditional knockout targeting vector construction. *Nucleic Acids Res.* **35**, e64 (2007).
- Warming, S., Rachel, R. A., Jenkins, N. A. & Copeland, N. G. *Zfp423* is required for normal cerebellar development. *Mol. Cell. Biol.* **26**, 6913–6922 (2006).
- Liu, H., Sadygov, R. G. & Yates, J. R. III. A model for random sampling and estimation of relative protein abundance in shotgun proteomics. *Anal. Chem.* **76**, 4193–4201 (2004).
- Gilchrist, A. *et al.* Quantitative proteomics analysis of the secretory pathway. *Cell* **127**, 1265–1281 (2006).
- Washburn, M. P., Wolters, D. & Yates, J. R. III. Large-scale analysis of the yeast proteome by multidimensional protein identification technology. *Nature Biotechnol.* **19**, 242–247 (2001).
- Pham, V. C. *et al.* Complementary proteomic tools for the dissection of apoptotic proteolysis events. *J. Proteome Res.* **11**, 2947–2954 (2012).
- Bakalarski, C. E. *et al.* The impact of peptide abundance and dynamic range on stable-isotope-based quantitative proteomic analyses. *J. Proteome Res.* **7**, 4756–4765 (2008).
- Slot, J. W. & Geuze, H. J. Cryosectioning and immunolabeling. *Nature Protocols* **2**, 2480–2491 (2007).



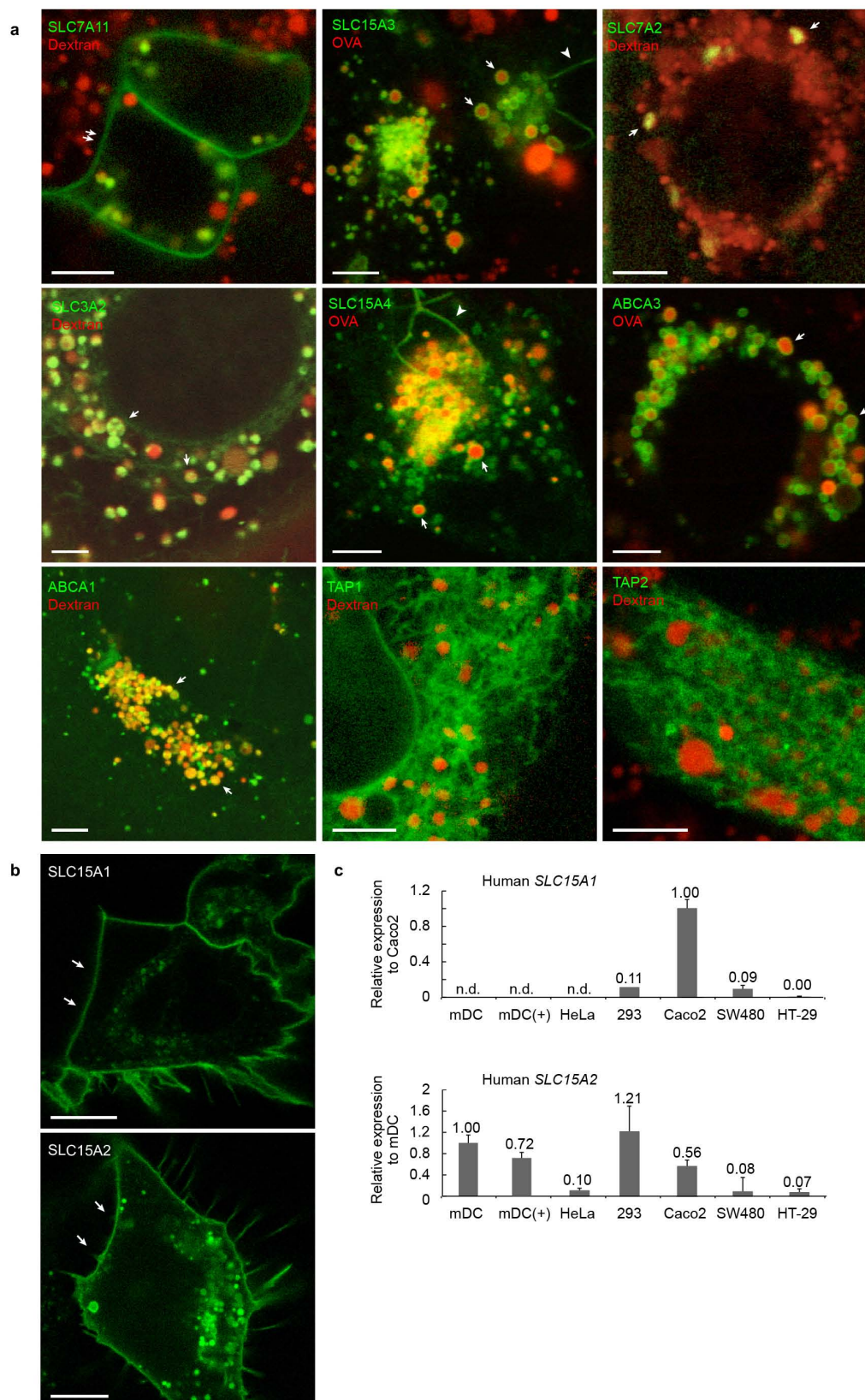
Extended Data Figure 1 | Phagocytosed *Salmonella* strongly activates NOD2. **a**, NOD2-transfected 293 cells were pre-incubated with or without cytochalasin D, then stimulated with soluble MDP. **b**, NOD2-transfected 293 cells were stimulated with wild-type or *invG* mutant of *S. typhimurium* harbouring a vector or a plasmid carrying *Yersinia* invasin gene (pInvasin) as in Fig. 1. **c**, **d**, 293 cells expressing LAMP1-GFP were incubated with wild-type (c) or *sifA* mutant (d) of *S. typhimurium* expressing tagBFP for 1 h, then cultured overnight in the presence of cultured for 24 h in the presence of 20 μ g mL⁻¹ gentamycin. The cells were observed under a fluorescence

microscope. Bars, 5 μ m. **e**, NOD2- or vector-transfected 293 cells were stimulated with wild-type or *sifA* mutant of *S. typhimurium* as in Fig. 1. **f**, NOD2-transfected 293 cells were stimulated with wild-type or *sifA* mutant of *S. typhimurium* harbouring vector or a plasmid carrying *SifA* gene as in Fig. 1. **g**, NOD2- or vector-transfected 293 cells were stimulated with wild-type, *sopE*, *sipA*, *sopE sipA* or *invG* mutant of *S. typhimurium*. **h**, NOD2-transfected cells were incubated with rhodamine-conjugated MDP-beads, then uptake by cells was quantified by FACS. One representative result of three independent experiments is shown. Mean of triplicates \pm s.d. **P* < 0.05 by *t*-test.



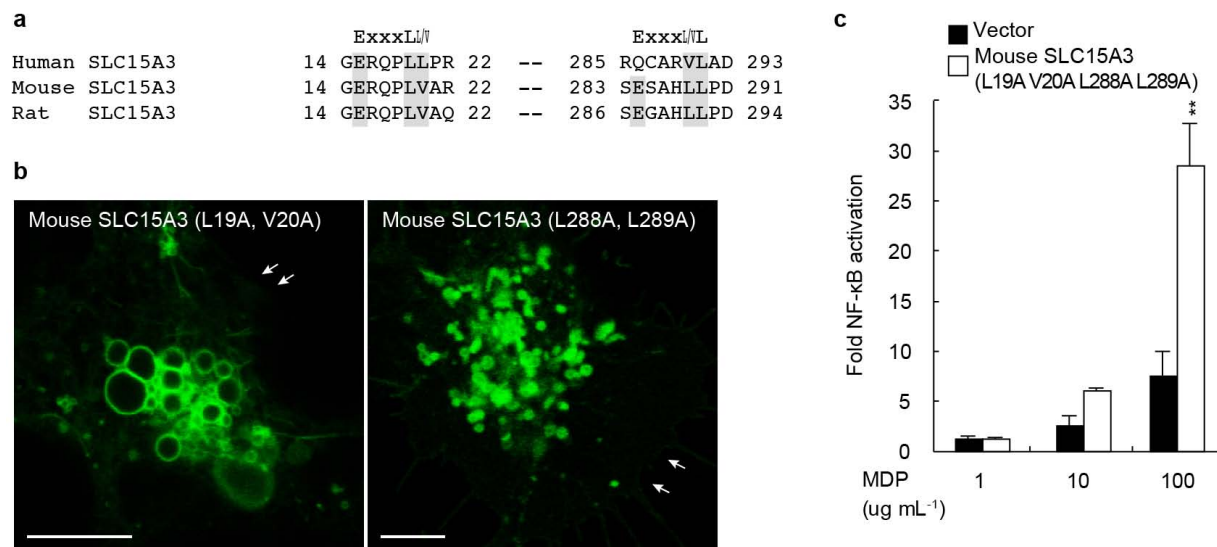
Extended Data Figure 2 | LAMP2 and MHCII are enriched in the endo-lysosome fraction of dendritic cells. **a, b,** Microscopic images of endo-lysosomes isolated from BMDCs stimulated without (**a**) or with (**b**) LPS. BMDCs were cultured with 200-nm diameter iron-oxide particles and ovalbumin (as an endocytic tracer), then incubated without (**a**) or with 30 ng ml⁻¹ LPS (**b**) for 3 h. The endo-lysosomes were isolated using magnetic separation from the cells lysed in isotonic buffer. Endocytosed ovalbumin-Alexa555 (OVA), MHCII-GFP, and bright field (BF) are shown. When dendritic cells are stimulated with LPS, isolated endo-lysosomes contained numerous tubular structures (arrows). Bars, 5 μ m. **c,** Total lysate, post-nuclear supernatant (PNS), flow through (unbound), and the fractions bound to column were blotted with antibodies against the marker proteins. Recoveries were calculated based on densitometric analysis using the ImageJ

(NIH). LAMP2 and MHCII are enriched 49- and 19-fold, respectively, in fraction 2 with 2% recovery as compared to the cell homogenate. Na⁺/K⁺-ATPase (plasma membrane), transferrin receptor (early endosomes), syntaxin-6 (trans-Golgi), and actin (cytosol) are excluded (0.12-, 0.057-, and 0.18-fold enrichment, respectively). **d,** SILAC mass spectrometry of endo-lysosomal proteins. Abundance ratios of LPS-stimulated endo-lysosomal proteins to unstimulated counterparts are plotted against their relative abundance. 1,436 proteins were identified in the endo-lysosome fractions, 8 and 45 proteins of which were ATP-binding cassette (ABC) transporters and solute carrier family transporters, respectively (shown in red). Eight (ABCA1, ABCA3, TAP1, TAP2, SLC3A2, SLC7A2, SLC7A11 and SLC15A3) were upregulated >twofold by LPS treatment.



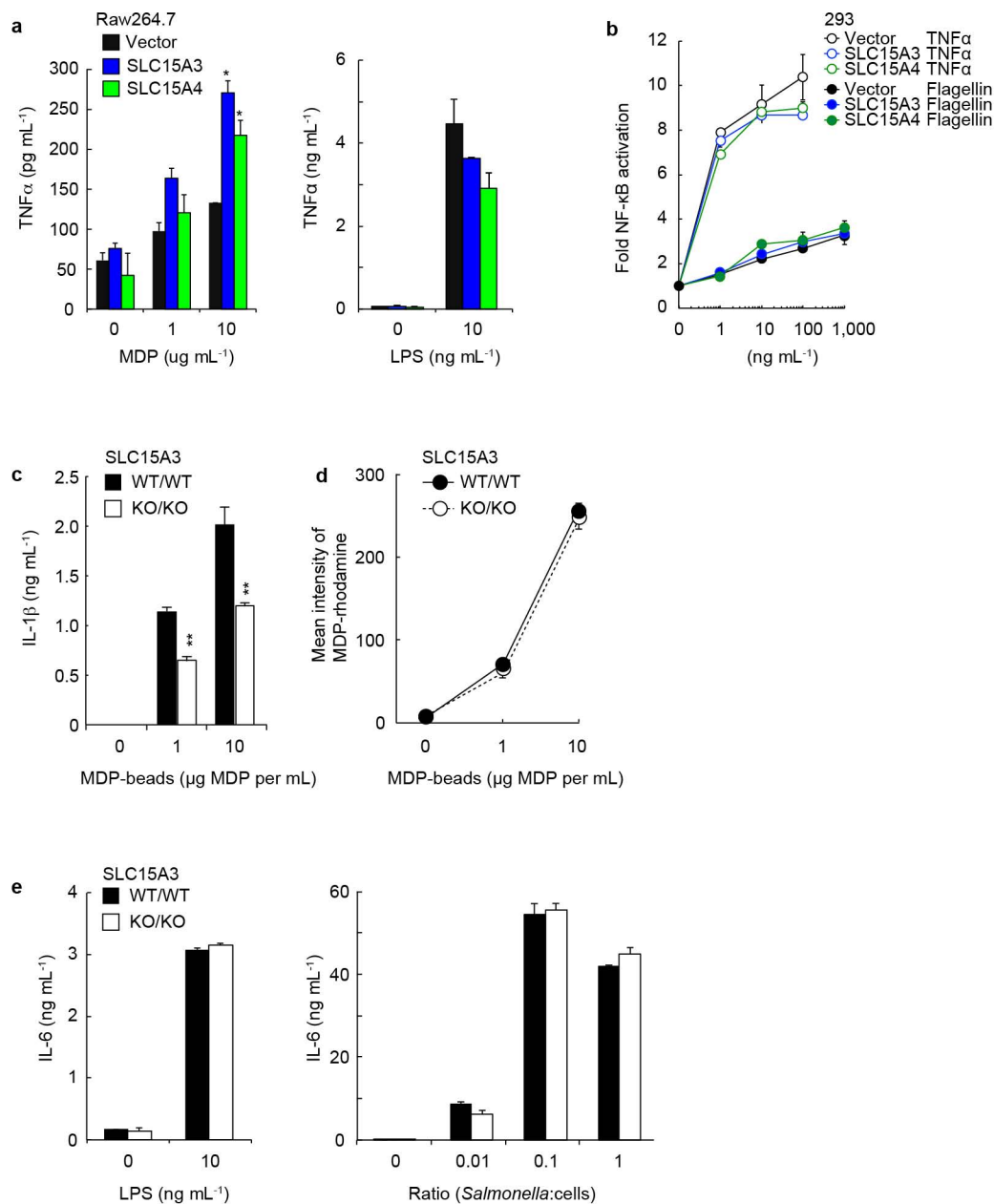
Extended Data Figure 3 | Localization of membrane transporters identified in the endo-lysosome fraction. **a**, Membrane transporters identified by mass spectrometry were expressed as GFP fusion proteins in 293 cells, then observed by fluorescence microscopy. Endo-lysosomes are visualized with ovalbumin–Alexa555 (OVA) or dextran–Alexa555. Single arrows indicate colocalization of membrane transporters labelled with endocytosed OVA or dextran. Tubular endo-lysosomes were observed only in cells expressing

SLC15A3 or SLC15A4 (arrow heads). Double arrows on cells expressing SLC7A11, ABCC4 or ABCC1 mark plasma membrane. **b**, SLC15A1 (mouse) or SLC15A2 (human) were expressed as GFP fusion proteins in 293 cells, then observed by fluorescence microscopy. Arrows mark plasma membrane. Bars, 5 μ m. **c**, Gene expression of *SLC15A1* and *SLC15A2* were quantified by RT–PCR. Means \pm s.d. (+) denotes LPS-stimulated dendritic cells.



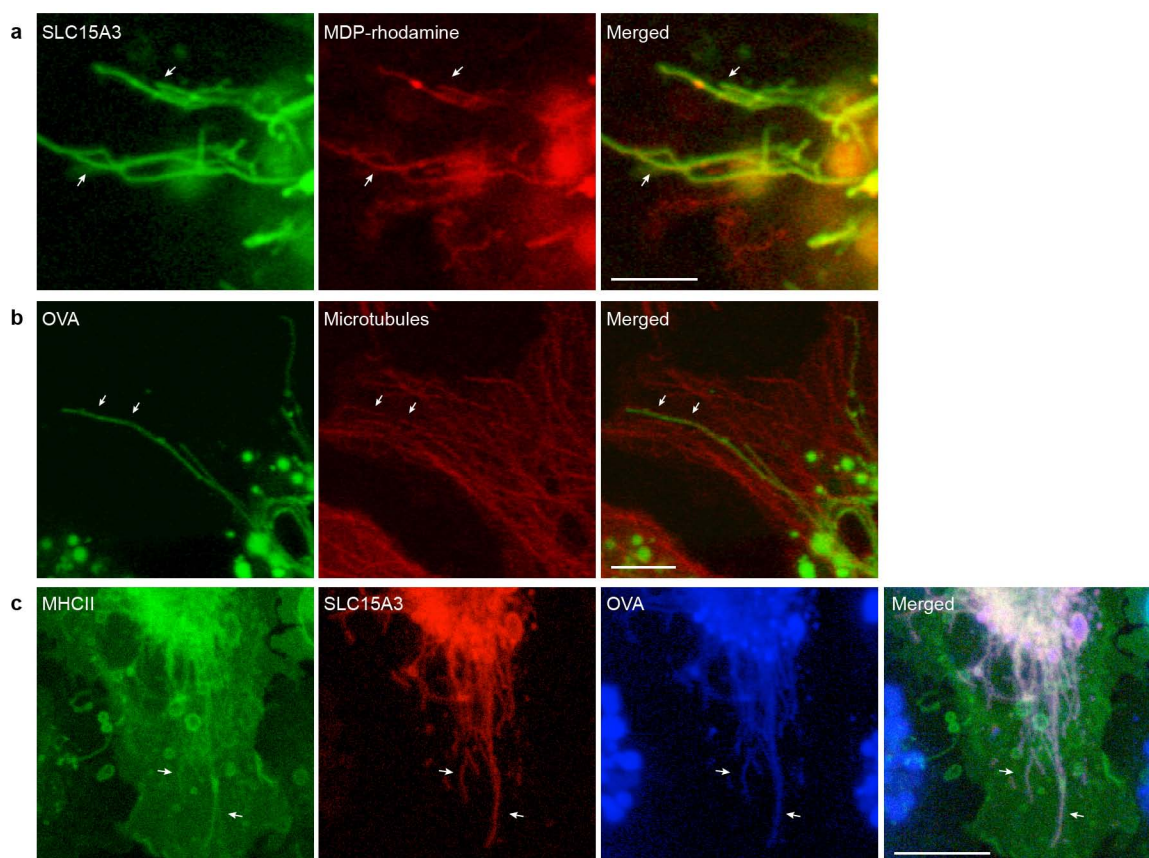
Extended Data Figure 4 | Two di-leucine motifs localize mouse SLC15A3 to endo-lysosomes. **a**, Di-leucine motifs in mammalian SLC15A3. Mouse and rat proteins have two di-leucine motifs (ExxxLL/V). **b**, Fluorescence microscopy of di-leucine mutants of mouse SLC15A3–GFP in 293 cells. Alanine substitution of either one in two motifs is insufficient to localize the protein to

plasma membrane. Cell surfaces are marked by arrows. Bars, 5 μm. **c**, 293 cells transfected with an NF-κB-luciferase reporter, NOD2, and either di-leucine mutant of mouse SLC15A3(L19A V20A L288A L289A) or empty vector, were stimulated with soluble MDP for 7 h. Means ± s.d. ***P* < 0.01.



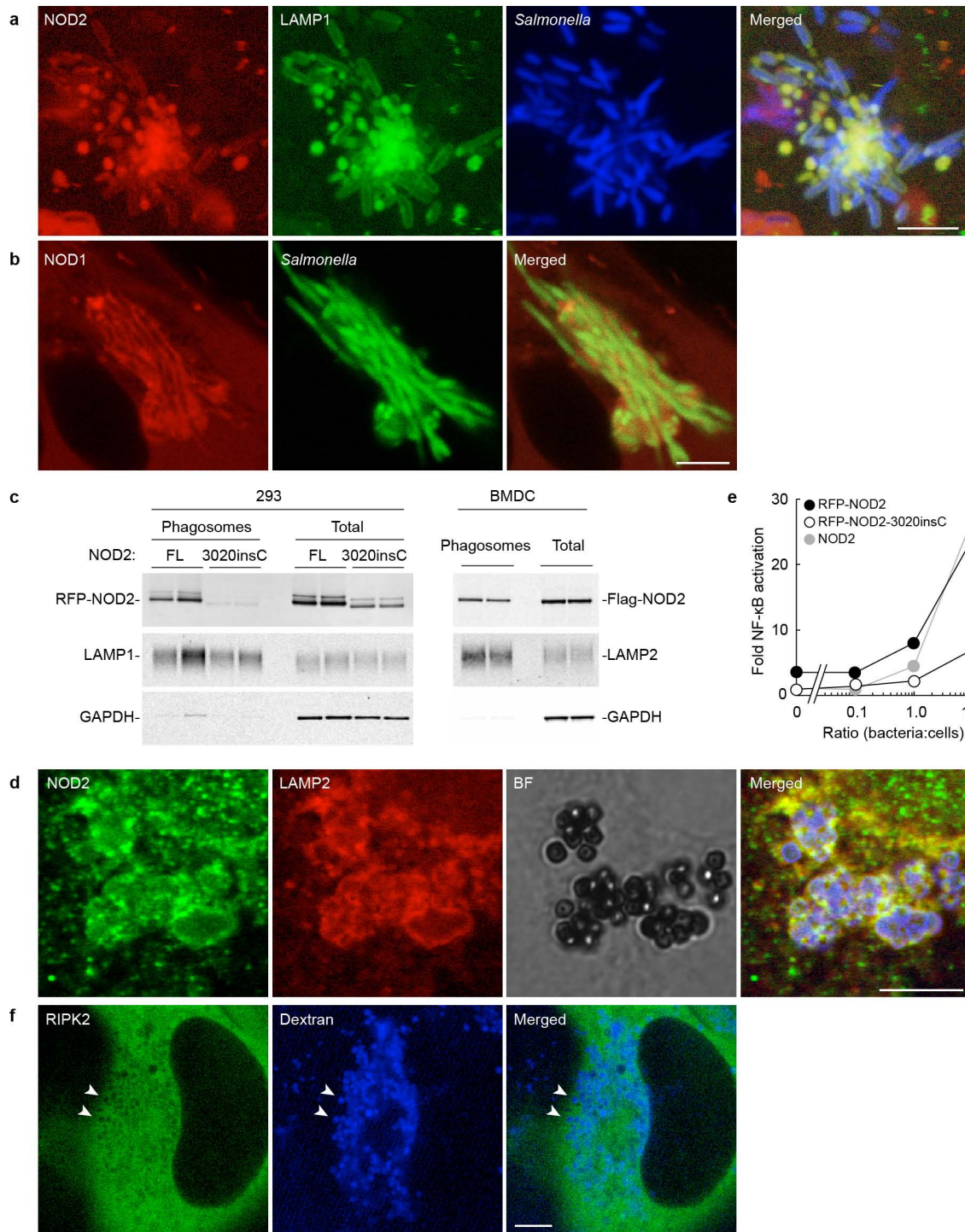
Extended Data Figure 5 | SLC15A3 and SLC15A4 specifically facilitate the signalling induced by MDP-beads. **a**, Raw 264.7 cells stably expressing SLC15A3 or SLC15A4 were stimulated with MDP (left) or LPS (right) for 24 h. TNF- α in medium was quantified by ELISA. **b**, 293 cells transfected with an NF- κ B-luciferase reporter, NOD2, and either SLC15A3, SLC15A4 or empty vector, were stimulated with *S. typhimurium* flagellin or human TNF- α for 7 h.

c, SLC15A3-deficient or wild-type BMDCs were stimulated with MDP-beads. IL-1 β in medium were quantified by ELISA. **d**, BMDCs were incubated with MDP-rhodamine-beads, then the uptake was quantified by FACS. **e**, SLC15A3-deficient or wild-type BMDCs were stimulated with LPS (left) or *S. typhimurium* (right) for 24 h. IL-6 in medium was quantified by ELISA. Means \pm s.d. * $P < 0.05$ and ** $P < 0.01$.



Extended Data Figure 6 | Endocytosed MDP are localized to endo-lysosomal tubules tracked along microtubules. **a**, 293 cells expressing SLC15A3–GFP were incubated with MDP–rhodamine for 6 h. After replacing with fresh medium without MDP–rhodamine, cells were observed by fluorescence microscopy. **b**, BMDCs that retrovirally express RFP–tubulin, were incubated with OVA–Alexa488 for 1 h, stimulated with LPS (30 ng ml^{-1})

for 3 h, then observed by fluorescence microscopy. Arrows indicate the endo-lysosomes extending along microtubules. **c**, MHC-II knock-in BMDCs that retrovirally express SLC15A3–GFP, were incubated with OVA–Alexa647 for 1 h, stimulated with LPS (30 ng ml^{-1}) for 3 h, then observed by fluorescence microscopy. Bars, $5 \mu\text{m}$.



Extended Data Figure 7 | NOD2 associates to *Salmonella*-containing phagosomes. **a**, 293 cells stably expressing RFP-NOD2 and LAMP1-GFP were incubated with *S. typhimurium* expressing tagBFP as in Fig. 4a. **b**, 293 cells expressing RFP-NOD1 were incubated with *S. typhimurium* expressing EGFP as in Fig. 4a, then observed by fluorescence microscopy. **c**, Left, 293 cells expressing RFP-NOD2 (FL, full length) or RFP-NOD2-3020insC were incubated with magnetic beads (1 µm), lysed in isotonic buffer, then the phagosomes were isolated using a magnetic column. Right, Phagosomes were isolated from BMDC expressing Flag-NOD2. RFP-NOD2, Flag-NOD2, LAMP1, LAMP2, and GAPDH in the phagosome fractions and total lysates were detected by immunoblotting of duplicated cell culture. NOD2 proteins in phagosomes were calculated by measuring the intensities of protein bands using ImageJ with following formula (1). GAPDH was used to estimate contamination of cytosol in the phagosome fraction. Approximately 20% of

NOD2 and 4% of NOD2-3020insC proteins are co-fractionated with LAMP1 or LAMP2 in the phagosome fractions. ($[\text{NOD2}]_{\text{fraction}}/[\text{NOD2}]_{\text{total lysate}} - [\text{GAPDH}]_{\text{fraction}}/[\text{GAPDH}]_{\text{total lysate}} / ([\text{LAMP1}]_{\text{fraction}}/[\text{LAMP1}]_{\text{total lysate}})$ (1). **d**, BMDCs retrovirally expressing Flag-NOD2 were incubated with MDP-beads for 24 h, fixed with 4% formaldehyde, permeabilized with 0.05% saponin, then stained with anti-Flag and anti-LAMP2 antibodies, followed by secondary antibodies conjugated with Alexa-dyes. The cells were observed under fluorescence microscopy. MDP-beads in the merged image are shown in blue. BF, bright field. **e**, 293 cells transfected with an NF-κB-luciferase reporter and either NOD2, RFP-NOD2, or RFP-NOD2-3020insC were incubated with *Salmonella* as in Fig. 1. The activation of NF-κB was quantified by the luciferase activity. **f**, Fluorescence microscopy of 293 cells expressing GFP-RIPK2. Endo-lysosomes are labelled with dextran-Alexa47. Bars, 5 µm.

Signal amplification and transduction in phytochrome photosensors

Heikki Takala^{1,2*}, Alexander Björling^{2*}, Oskar Berntsson², Heli Lehtivuori¹, Stephan Niebling², Maria Hoernke², Irina Kosheleva³, Robert Henning³, Andreas Menzel⁴, Janne A. Ihalainen¹ & Sebastian Westenhoff²

Sensory proteins must relay structural signals from the sensory site over large distances to regulatory output domains. Phytochromes are a major family of red-light-sensing kinases that control diverse cellular functions in plants, bacteria and fungi^{1–9}. Bacterial phytochromes consist of a photosensory core and a carboxy-terminal regulatory domain^{10,11}. Structures of photosensory cores are reported in the resting state^{12–18} and conformational responses to light activation have been proposed in the vicinity of the chromophore^{19–23}. However, the structure of the signalling state and the mechanism of downstream signal relay through the photosensory core remain elusive. Here we report crystal and solution structures of the resting and activated states of the photosensory core of the bacteriophytochrome from *Deinococcus radiodurans*. The structures show an open and closed form of the dimeric protein for the activated and resting states, respectively. This nanometre-scale rearrangement is controlled by refolding of an evolutionarily conserved ‘tongue’, which is in contact with the chromophore. The findings reveal an unusual mechanism in which atomic-scale conformational changes around the chromophore are first amplified into an ångström-scale distance change in the tongue, and further grow into a nanometre-scale conformational signal. The structural mechanism is a blueprint for understanding how phytochromes connect to the cellular signalling network.

The most common domain architecture of the photosensory core of phytochromes in bacteria, plants and fungi is PAS-GAF-PHY (Per/Arndt/Sim-cGMP phosphodiesterase/adenyl cyclase/FhlA-phytochrome specific)¹⁰. We investigate the PAS-GAF-PHY fragment of the bacteriophytochrome from *D. radiodurans*. Figure 1a shows that this protein fragment shares the prototypical response of phytochromes to red/far-red illumination. Its biliverdin chromophore relaxes into the Pr (red-absorbing) state when kept in the dark or after exposure to far-red light. The majority of molecules transform into the Pfr (far-red-absorbing) state after red-light exposure. The dark relaxation from Pfr to Pr occurs in a matter of days, and the optical response of PAS-GAF-PHY is very similar to the full-length protein²⁴. For the PHY-less PAS-GAF construct, the absorption spectra (Fig. 1a) and dark reversion within minutes indicate an incomplete photocycle²⁴.

Time-resolved solution X-ray scattering reports on structural changes in protein reactions^{25,26}. We used it here to characterize the structural dynamics of the phytochrome photocycle. In a first experiment, the phytochrome fragments were repeatedly switched between the Pr and Pfr states using laser flashes of 10-ms duration at 671 nm (Pr → Pfr) and 750 nm (Pfr → Pr). The X-ray solution scattering was recorded in between the laser flashes for approximately 1 s (see Supplementary Information for details). The light-induced change in scattering, $\Delta S(q)$, encodes the structural change of the protein during the Pr → Pfr transition. For the complete photosensory core (PAS-GAF-PHY) we find a large oscillating difference scattering signal at low angles, $q < 2 \text{ nm}^{-1}$ (Fig. 1b, red line), indicating large structural changes on the nanometre scale. Oscillations at higher q , which report on structural changes on smaller

length scales, were also observed (Extended Data Fig. 1e), but are not discussed further here. The laser-induced ΔS of the PAS-GAF-PHY construct was reproduced by a standard small-angle X-ray scattering (SAXS) experiment with pre-illuminated samples (Fig. 1b, blue line). Time-resolved difference X-ray scattering data covering the micro- and milli-second time scales also reproduce these features when photoconversion is complete (delay time, 30 ms; Fig. 1b, black line). These data further establish that the discussed low- q features grow in with a time-constant of 4.3 ms (Fig. 1b and Extended Data Fig. 1d), which coincides with the formation of the Pfr state as measured by optical absorption of the chromophore (1.0 ms, Extended Data Fig. 1d).

When performing the same experiments on the PHY-less chromophore-binding (PAS-GAF), the oscillatory signal at low q is absent (Fig. 1c). The observed oscillations in the PAS-GAF domains are much smaller in amplitude and confined to higher q ranges (Fig. 1c and Extended Data Fig. 1e). These data show that quaternary structural changes occur only when the PHY domain is present. In the wild-type protein the PHY domain connects the chromophore binding domains PAS-GAF with the output kinase, making it likely that the detected rearrangement is functionally relevant.

To examine the nature of these conformational changes and to find out how they arise, we crystallized the photosensory core in the dark (referred to as dark) and under periodic illumination at 655 nm (referred to as illuminated), which transforms a large fraction of the protein molecules into the Pfr state (see Fig. 1a and Extended Data figures for details). Absorption spectra of the dark crystals indicate the Pr state, whereas the illuminated crystals are Pfr-enriched (Extended Data Fig. 3c). Omit maps indicate a change in the conformation of the chromophore between the two crystal forms (Extended Data Fig. 3b). Although the biliverdin conformation cannot be fit unambiguously to the electron density in the illuminated structure, the electron density of the rest of the protein is homogeneous (Extended Data Fig. 2b).

Our dark crystal structure was modelled from data up to 3.80 Å resolution (Fig. 2a). The observed domain arrangement is similar to previously reported structures of bacterial PAS-GAF and PAS-GAF-PHY protein fragments^{12–15,17}, but the dimer arrangement differs from some of these structures^{13,17}. The illuminated structure, modelled against data cut at 3.24 Å , shows marked differences compared to the dark structure (Fig. 2a). First, the dimer adopts an open Y-like shape of the PHY domains. This is in contrast to all reported resting state structures, where the dimer is closed (Fig. 2a)^{12,13,17,20}. Second, the tongue of the PHY domains (residues 446–477) appears as an α -helix and a loop in the illuminated model and is a β -sheet in the dark model (Fig. 2b and Extended Data Fig. 2). These two folds have been observed separately in non-canonical bacterial^{12,20} and prototypical cyanobacterial phytochromes^{13,17}, respectively (Extended Data Fig. 4). Our structures establish that the refolding of the tongue occurs within the same bacterial phytochrome, and suggest that this refolding is associated with opening of the dimer. The change in fold causes the length of the tongue to vary between the

¹Nanoscience Center, Department of Biological and Environmental Science, University of Jyväskylä, 40014 Jyväskylä, Finland. ²Department of Chemistry and Molecular Biology, University of Gothenburg, 40530 Gothenburg, Sweden. ³Center for Advanced Radiation Sources, The University of Chicago, Illinois 60637, USA. ⁴Paul Scherrer Institut, 5232 Villigen PSI, Switzerland.

*These authors contributed equally to this work.

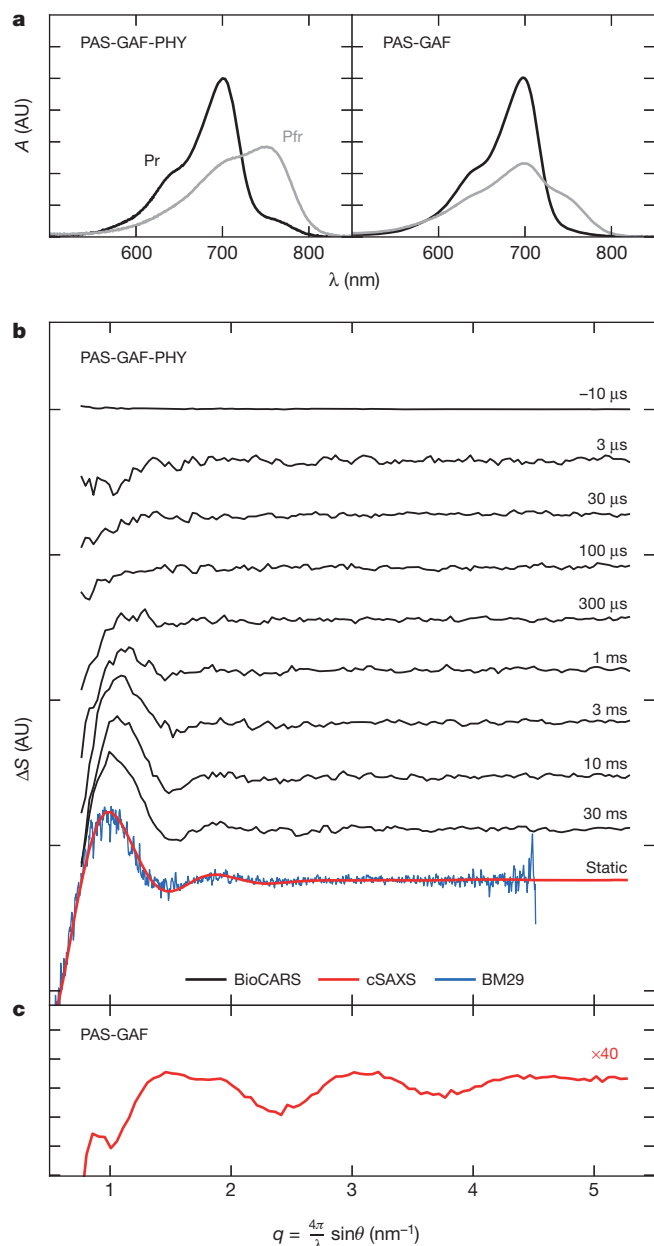


Figure 1 | Time-resolved solution X-ray scattering of the PAS-GAF and PAS-GAF-PHY fragments from *D. radiodurans*. **a**, Absorption (A) spectra of protein fragments after illumination with far-red (780 nm) and red (655 nm) light, labelled Pr and Pfr respectively. AU, arbitrary units. **b**, **c**, Solution X-ray scattering data from the PAS-GAF-PHY and PAS-GAF fragments shown on the same scale. Time-resolved data (black, BioCARS), direct static data collected by laser-induced population switching (red, cSAXS) and indirect static data from a standard SAXS experiment with pre-illuminated samples (blue, BM29) is shown. BioCARS, cSAXS and BM29 refer to synchrotron endstations (see Methods). ΔS is the difference in scattered X-ray intensity caused by illumination at 671 nm. $q = 4\pi/\lambda \times \sin \theta$ at wavelength λ and scattering angle 2θ .

dark and illuminated structures (Fig. 2b), which is key for understanding the role of the tongue in signal relay.

Before detailing this mechanism, it is important to test how well the crystal models represent the solution structures of the Pr and Pfr states. To this end, we performed structural refinement against the solution difference X-ray scattering data (Fig. 1), which is a sensitive indicator of conformational change. It has the advantage over absolute scattering profiles that experimental error and uncertainties arising from the choice of method for scattering calculations largely cancel out (Extended Data Fig. 5c).

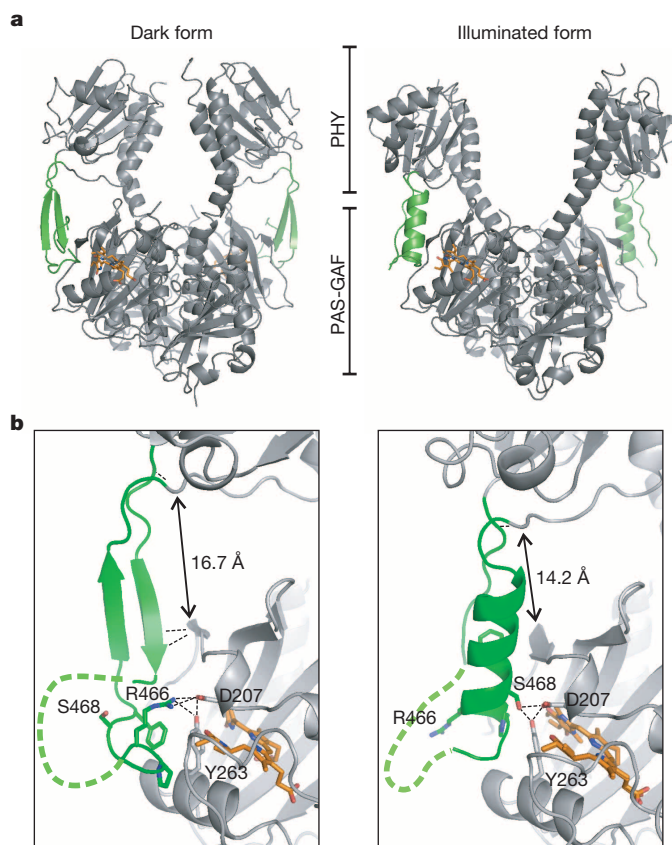


Figure 2 | Dark and illuminated crystal structures of the PAS-GAF-PHY fragment from *D. radiodurans*. **a**, Crystal structures of the PAS-GAF-PHY dimer in the dark and illuminated forms. The tongue of the PHY domain (green) changes fold and the dimer opens up in the illuminated state. The biliverdin chromophore is shown in orange. **b**, Fold and interactions of the PHY tongue. The β -sheet (dark) coordinates to Asp 207 and Tyr 263 via Arg 466, whereas the α -helix (illuminated) coordinates via Ser 468, both of which are part of the conserved $^{465}\text{P-R-X-S-F}^{469}$ motif. The named residues are shown as sticks. The β -sheet (dark) is further held by hydrogen-bonding interactions between the amide groups of Ala 450, Gly 452 and Arg 202. The change in PHY domain conformation leads to a shortening of the tongue by 2.5 Å as measured between GAF (Arg 202) and PHY (Tyr 479) domains (arrows). A backbone interaction close to the PHY domain between Leu 445 and Tyr 479, shared by both crystal structures, is also indicated. The green dashed lines indicate regions that are not modelled or not shown for clarity.

We generated candidate solution structures by running two sets of molecular dynamics simulations starting from the dark and illuminated crystal structures (see Supplementary Information for details). These are referred to as Pr and Pfr trajectories as they aim to model solution scattering data representing these states. Snapshots were recorded every 50 ps and the solution X-ray scattering was calculated from each³⁰. All pairs of Pr/Pfr candidate structures were then compared to the static difference X-ray scattering (Fig. 1b and Extended Data Fig. 7, see Supplementary Information for details). We selected the 100 Pr/Pfr pairs with best agreement to the data and consider the participating structures to be representative of the protein in solution. Considering all pairwise differences between these structures, a consistent set of 747 curves was generated and evaluated (Fig. 3a). The agreement of the model and experiment is excellent, and considerable improvement is made compared to the crystal structures.

To verify the choice of solution structures on the basis of difference scattering, these structures were cross-validated against absolute, population-corrected SAXS data as described in Supplementary Information (see also Extended Data Fig. 5). Figure 3b shows that the agreement with absolute SAXS data is strongly correlated with the separation of the PHY domains, and that the solution structures, proposed purely on

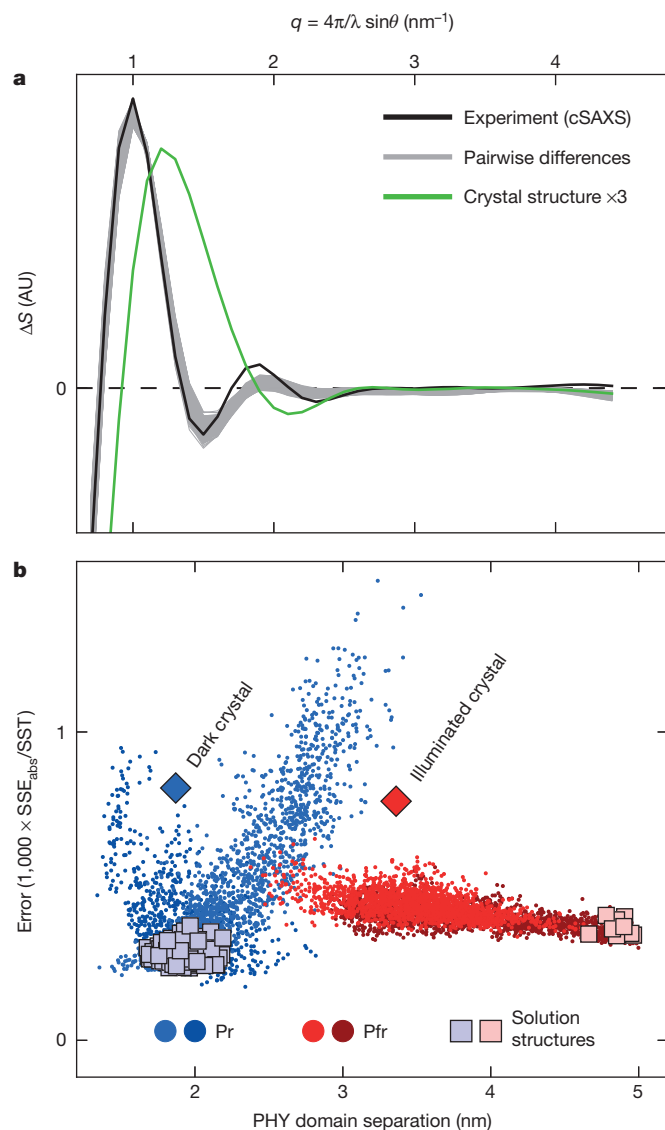


Figure 3 | Refinement of solution structures against difference SAXS data.

a, Calculated differences X-ray scattering between the proposed solution structures (Pfr – Pr, grey) agree with experimental data (black). As a safeguard against overfitting, the grey curves show all 747 differences between solution structures, not just the 100 best curves on which the pairs were selected (see text for details). The agreement with experiment is considerably improved for the solution structures, compared to the difference scattering calculated from the two crystal forms (green). **b**, Validation of the obtained solution structures (squares) against absolute X-ray scattering. The vertical axis shows the sum-squared error of each structure (SSE_{abs}) compared to the total sum of squares about the mean (SST). All molecular dynamics snapshots (dots) and crystal structures (diamonds) are scored against absolute SAXS data (BM29) after correction for the mixed Pr/Pfr populations (see Extended Data Fig. 5d). The two shades of red and blue correspond to different simulation conditions as detailed in Supplementary Information. The PHY domain separation is measured as the distance between the centres of mass of the two C-terminal helices (residues 484–503).

difference scattering, cluster in the low-error parts of these correlations. The agreement of the solution structures to the solution scattering is considerably improved compared to the crystal structures as can be seen in Fig. 3b. Representative structural models for Pr and Pfr states in solution are depicted in Fig. 4.

The solution-structural models show that the PHY domains of opposing monomers come fairly close to each other in the Pr state but move apart by ~3 nm during the Pr → Pfr transition (Fig. 4). The Y shape assumed by the solution structures in the Pfr state is in qualitative

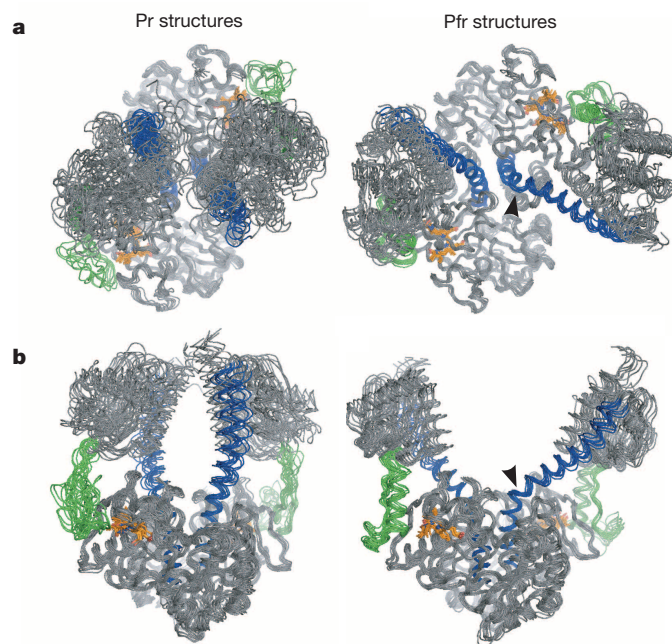


Figure 4 | Proposed solution structures of the bacterial phytochrome from *D. radiodurans*.

a, b, Representative solution structures for the Pr and Pfr states of the photosensory core, identified from solution X-ray scattering experiments. Nine Pfr and ten Pr structures are presented viewed along (**a**) and perpendicular (**b**) to the dimer symmetry axis. The long scaffolding helix is highlighted in blue, the PHY tongue in green, and the biliverdin chromophore in orange. The PHY domain separation differs by about 3 nm between the Pr and Pfr structures as shown in Fig. 3b. The hinge region at the scaffolding helix (residues Val 318 and Lys 319) in Pfr is indicated with black arrowheads.

agreement with a low-resolution envelope of a *Rhodospseudomonas palustris* phytochrome determined by SAXS²⁷. Our crystal structures capture the essence of this quaternary conformational change, albeit with differences in the amplitude and details of the motion. These differences are probably caused by crystal contacts (Extended Data Fig. 6). Nevertheless, the overall agreement with the solution structures lends strong support to the idea that the refolding of the tongue observed in the crystal structures is intrinsic to the structural rearrangement between Pr and Pfr states.

From the crystal and solution structures the following mechanism for signal transduction through the photosensory core emerges. It is known that the rotation of the biliverdin D-ring causes atomic rearrangements in the chromophore binding pocket, including displacement of Asp 207 and Tyr 263 (refs 19–22). Our crystal structures suggest that this controls the fold of the tongue of the PHY domain. Importantly, the tongue is shortened as a result of the refolding and the distance between the GAF and PHY domain is reduced by 2.5 Å (Fig. 2b) upon red-light illumination. As a consequence of this, and demonstrated by our Pr and Pfr solution structures, the dimer opens up between the PHY domains by several nanometres (Fig. 4).

Three factors are essential for the proposed structural mechanism. First, the PAS, GAF and PHY domains, which are connected by the tongue, are known to be very rigid internally such that they cannot deform to absorb the length variation of the tongue¹¹. Second, the tongue and its junctions to PHY and GAF have to be rigid, especially in Pr, in which the PHY and GAF domains are pushed away from each other. In our dark structures the direction of the tongue is rigidly fixed with respect to the GAF domains by the conserved interaction Arg 466 to Asp 207 and Tyr 263 and additional backbone hydrogen bonding (Fig. 2b)¹³. In the illuminated structure the tongue binds to GAF through interaction of Ser 468 with Asp 207 and Tyr 263 (ref. 12), and through hydrophobic interactions, among others by Tyr 472 (Extended Data Fig. 4b)¹⁷. The

tongue is also locked close to the PHY domain by additional backbone hydrogen bonding. Third, the PAS-GAF and PHY domains are connected by the unusually long scaffolding helix, which is necessary to redirect the shortening of the tongue into bending of the monomer. Indeed, our simulations reveal a hinge in the scaffolding helix (Fig. 4b). All phytochrome structures published to date support these three requirements^{12–18}.

In bacterial and fungal phytochromes, signal output is through C-terminal histidine kinase domains that autophosphorylate and further phosphotransfer to a response regulator^{3,7}. In plant phytochromes, two additional PAS domains are included in the C-terminal regulatory region and a more complex pattern of functions has to be controlled, such as serine/threonine kinase activity⁴, and affinity to interaction partners^{9,10}. In all cases, the output activity is probably controlled by a structural change in the photosensory core. The photosensory core and the key amino acid sequence in the tongue region ⁴⁶⁵P-R-X-S-F⁴⁶⁹ (where X denotes any residue) are highly conserved over the whole phytochrome superfamily (Extended Data Fig. 2c). Here we suggest a mechanism in which the fold of the PHY-tongue controls the bending of the monomer. For the isolated photosensory unit in this study, this results in marked opening of the dimer in Pfr (Figs 2 and 4). In full-length phytochromes it becomes conceivable that the conformations of output domains are modified. Alternatively, previously buried patches of the protein could become accessible to interaction partners, or monomers could rearrange with respect to each other. It will be intriguing to see the emergence of these mechanisms, which become deductible with experimental approaches similar to the one presented here.

METHODS SUMMARY

The PAS-GAF-PHY and PAS-GAF fragments from *D. radiodurans* were expressed in the *Escherichia coli* strain BL21 (DE3) and purified by affinity and size-exclusion chromatography. Crystallographic data was collected at beamline ID23-1 of the European Synchrotron Radiation Facility (ESRF) (see Extended Data Table 1). Time-resolved X-ray scattering with millisecond time resolution were recorded at beamline cSAXS of the Swiss Light Source²⁸. SAXS measurements were performed at beamline BM29 of the ESRF and analysed as summarized in Extended Data Table 2a. Time-resolved X-ray scattering data in the micro- and millisecond ranges were collected at beamline ID-14-B, BioCARS, of the Advanced Photon Source at Argonne National Laboratory. All solution-scattering sample details are summarized in Extended Data Table 2b. Molecular dynamics simulations (GROMACS 4.5.5)²⁹ were used to generate trial solution structures and theoretical scattering curves were evaluated using Zernike expansion as implemented in SASTBX³⁰. The structures were scored against experimental scattering data as detailed in Supplementary Information.

Online Content Any additional Methods, Extended Data display items and Source Data are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 23 December 2013; accepted 7 April 2014.

Published online 30 April 2014.

- Butler, W. L., Norris, K. H., Siegelman, H. W. & Hendricks, S. B. Detection, assay, and preliminary purification of the pigment controlling photoresponsive development of plants. *Proc. Natl Acad. Sci. USA* **45**, 1703–1708 (1959).
- Kehoe, D. M. & Grossman, A. R. Similarity of a chromatic adaptation sensor to phytochrome and ethylene receptors. *Science* **273**, 1409–1412 (1996).
- Yeh, K. C., Wu, S. H., Murphy, J. T. & Lagarias, J. C. A cyanobacterial phytochrome two-component light sensory system. *Science* **277**, 1505–1508 (1997).
- Yeh, K.-C. & Lagarias, J. C. Eukaryotic phytochromes: light-regulated serine/threonine protein kinases with histidine kinase ancestry. *Proc. Natl Acad. Sci. USA* **95**, 13976–13981 (1998).
- Jiang, Z. *et al.* Bacterial photoreceptor with similarity to photoactive yellow protein and plant phytochromes. *Science* **285**, 406–409 (1999).
- Hughes, J. *et al.* A prokaryotic phytochrome. *Nature* **386**, 663 (1997).
- Bhoo, S. H., Davis, S. J., Walker, J., Karniol, B. & Vierstra, R. D. Bacteriophytochromes are photochromic histidine kinases using a biliverdin chromophore. *Nature* **414**, 776–779 (2001).
- Blumenstein, A. *et al.* The *Aspergillus nidulans* phytochrome FphA represses sexual development in red light. *Curr. Biol.* **15**, 1833–1838 (2005).
- Ni, M., Tepperman, J. M. & Quail, P. H. Binding of phytochrome B to its nuclear signalling partner PIF3 is reversibly induced by light. *Nature* **400**, 781–784 (1999).
- Rockwell, N. C., Su, Y. S. & Lagarias, J. C. Phytochrome structure and signaling mechanisms. *Annu. Rev. Plant Biol.* **57**, 837–858 (2006).

- Möglich, A., Yang, X. J., Ayers, R. A. & Moffat, K. Structure and function of plant photoreceptors. *Annu. Rev. Plant Biol.* **61**, 21–47 (2010).
- Yang, X., Kuk, J. & Moffat, K. Crystal structure of *Pseudomonas aeruginosa* bacteriophytochrome: photoconversion and signal transduction. *Proc. Natl Acad. Sci. USA* **105**, 14715–14720 (2008).
- Essen, L. O., Mailliet, J. & Hughes, J. The structure of a complete phytochrome sensory module in the Pr ground state. *Proc. Natl Acad. Sci. USA* **105**, 14709–14714 (2008).
- Wagner, J. R., Brunzelle, J. S., Forest, K. T. & Vierstra, R. D. A light-sensing knot revealed by the structure of the chromophore-binding domain of phytochrome. *Nature* **438**, 325–331 (2005).
- Li, H., Zhang, J. R., Vierstra, R. D. & Li, H. L. Quaternary organization of a phytochrome dimer as revealed by cryoelectron microscopy. *Proc. Natl Acad. Sci. USA* **107**, 10872–10877 (2010).
- Cornilescu, G., Uliasz, A. T., Cornilescu, C. C., Markley, J. L. & Vierstra, R. D. Solution structure of a cyanobacterial phytochrome GAF domain in the red-light-absorbing ground state. *J. Mol. Biol.* **383**, 403–413 (2008).
- Anders, K., Daminelli-Widany, G., Mroginiski, M. A., von Stetten, D. & Essen, L. O. Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling. *J. Biol. Chem.* **288**, 35714–35725 (2013).
- Narikawa, R. *et al.* Structures of cyanobacteriochromes from phototaxis regulators AnPixJ and TePixJ reveal general and specific photoconversion mechanism. *Proc. Natl Acad. Sci. USA* **110**, 918–923 (2013).
- Yang, X., Kuk, J. & Moffat, K. Conformational differences between the Pfr and Pr states in *Pseudomonas aeruginosa* bacteriophytochrome. *Proc. Natl Acad. Sci. USA* **106**, 15639–15644 (2009).
- Yang, X., Ren, Z., Kuk, J. & Moffat, K. Temperature-scan cryocrystallography reveals reaction intermediates in bacteriophytochrome. *Nature* **479**, 428–432 (2011).
- Song, C. *et al.* Two ground state isoforms and a chromophore D-ring photoflip triggering extensive intramolecular changes in a canonical phytochrome. *Proc. Natl Acad. Sci. USA* **108**, 3842–3847 (2011).
- Burgie, E. S., Walker, J. M., Phillips, G. N. & Vierstra, R. D. A Photo-labile thioether linkage to phycoviolobilin provides the foundation for the blue/green photocycles in DXCF-cyanobacteriochromes. *Structure* **21**, 88–97 (2013).
- Uliasz, A. T. *et al.* Structural basis for the photoconversion of a phytochrome to the activated Pfr form. *Nature* **463**, 250–254 (2010).
- Wagner, J. R. *et al.* Mutational analysis of *Deinococcus radiodurans* bacteriophytochrome reveals key amino acids necessary for the photochromicity and proton exchange cycle of phytochromes. *J. Biol. Chem.* **283**, 12212–12226 (2008).
- Camarata, M. *et al.* Tracking the structural dynamics of proteins in solution using time-resolved wide-angle X-ray scattering. *Nature Methods* **5**, 881–886 (2008).
- Andersson, M. *et al.* Structural dynamics of light-driven proton pumps. *Structure* **17**, 1265–1275 (2009).
- Evans, K., Grossmann, J. G., Fordham-Skelton, A. P. & Papiz, M. Z. Small-angle X-ray scattering reveals the solution structure of a bacteriophytochrome in the catalytically active Pr state. *J. Mol. Biol.* **364**, 655–666 (2006).
- Westenhoff, S. *et al.* Rapid readout detector captures protein time-resolved WAXS. *Nature Methods* **7**, 775–776 (2010).
- Pronk, S. *et al.* GROMACS 4.5: a high-throughput and highly parallel open source molecular simulation toolkit. *Bioinformatics* **29**, 845–854 (2013).
- Liu, H. G., Hexemer, A. & Zwart, P. H. The Small Angle Scattering ToolBox (SASTBX): an open-source software for biomolecular small-angle scattering. *J. Appl. Crystallogr.* **45**, 587–593 (2012).

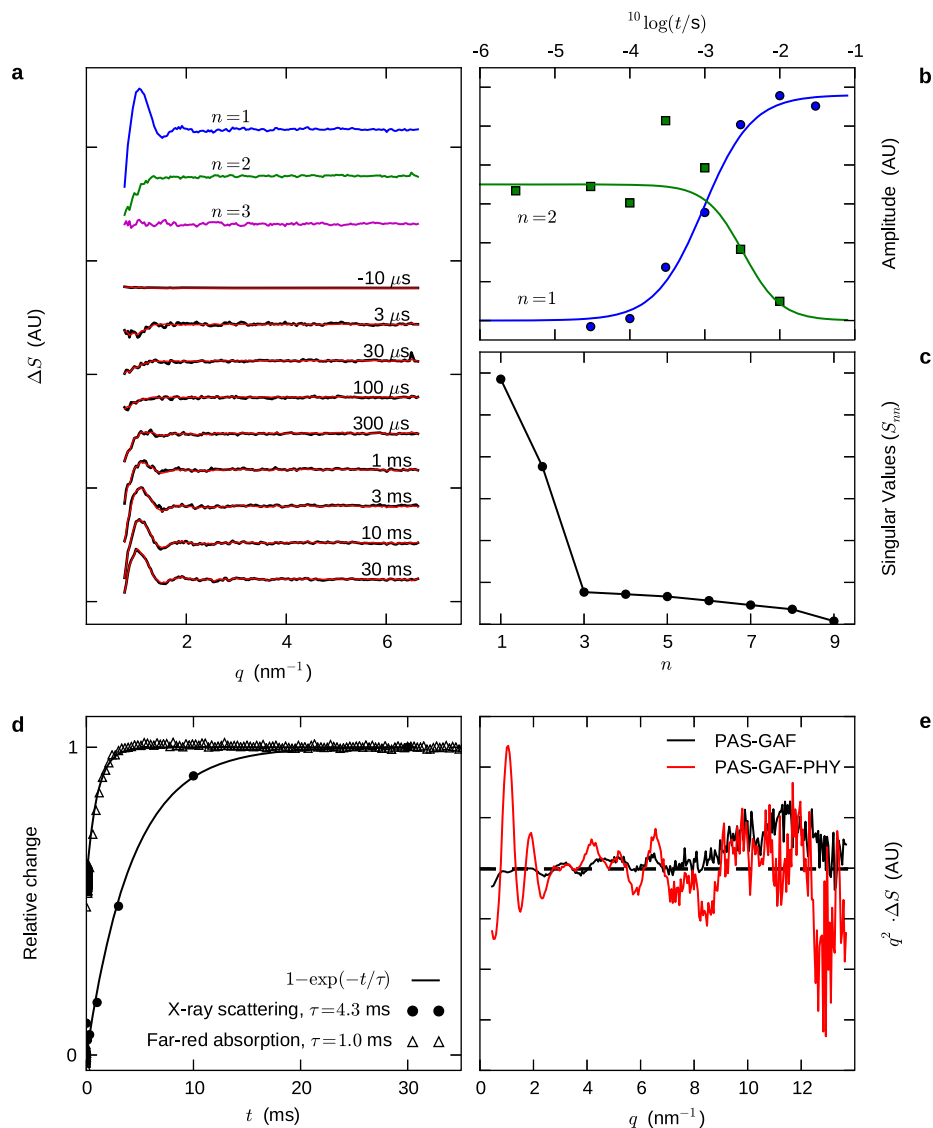
Supplementary Information is available in the online version of the paper.

Acknowledgements We acknowledge beamline access at BM29 and ID23-1 at the ESRF, cSAXS at the Swiss Light Source (funded by the European Community's FP7 grant agreement no. 312284 CALIPSO), and BioCARS at the Advanced Photon Source, Argonne National Laboratory (supported by National Institutes of Health, National Institute of General Medical Sciences grant 1R24GM111072 and in part through collaboration with P. Anfinrud). K. Forest and R. Vierstra are acknowledged for the donation of the plasmids. J. Yläne and his group are acknowledged for advice on crystallography, and H. Häkkinen and A. Liukkonen for their assistance. We thank M.-A. Mroginiski for providing force-field parameter files for the biliverdin. We acknowledge V. Chukharev and N. V. Tkachenko for the facilities for transient absorption spectroscopy measurements. We also acknowledge grants to S.W. by the Foundation of Strategic Research, Sweden, and the Swedish and European Research Councils, agreement number 279944, support to J.A.I. by Finnish Academy grant 138063, and to H.T. by the Finnish Cultural Foundation grant 0131067.

Author Contributions J.A.I., S.W., A.B. and H.T. conceived the experiments; H.T. prepared the samples, crystallized the protein, solved the structure and measured crystal spectra; A.B., H.T., S.W., J.A.I., S.N., H.L., M.H., A.M., R.H. and I.K. performed the solution X-ray scattering experiments; H.L. performed the spectroscopic measurements in solution; A.B. performed the molecular dynamics simulations; A.B., H.T., O.B. and S.N. analysed data; S.W. and J.A.I. supervised all parts of the project; and S.W., J.A.I., H.T. and A.B. wrote the paper with input from all authors.

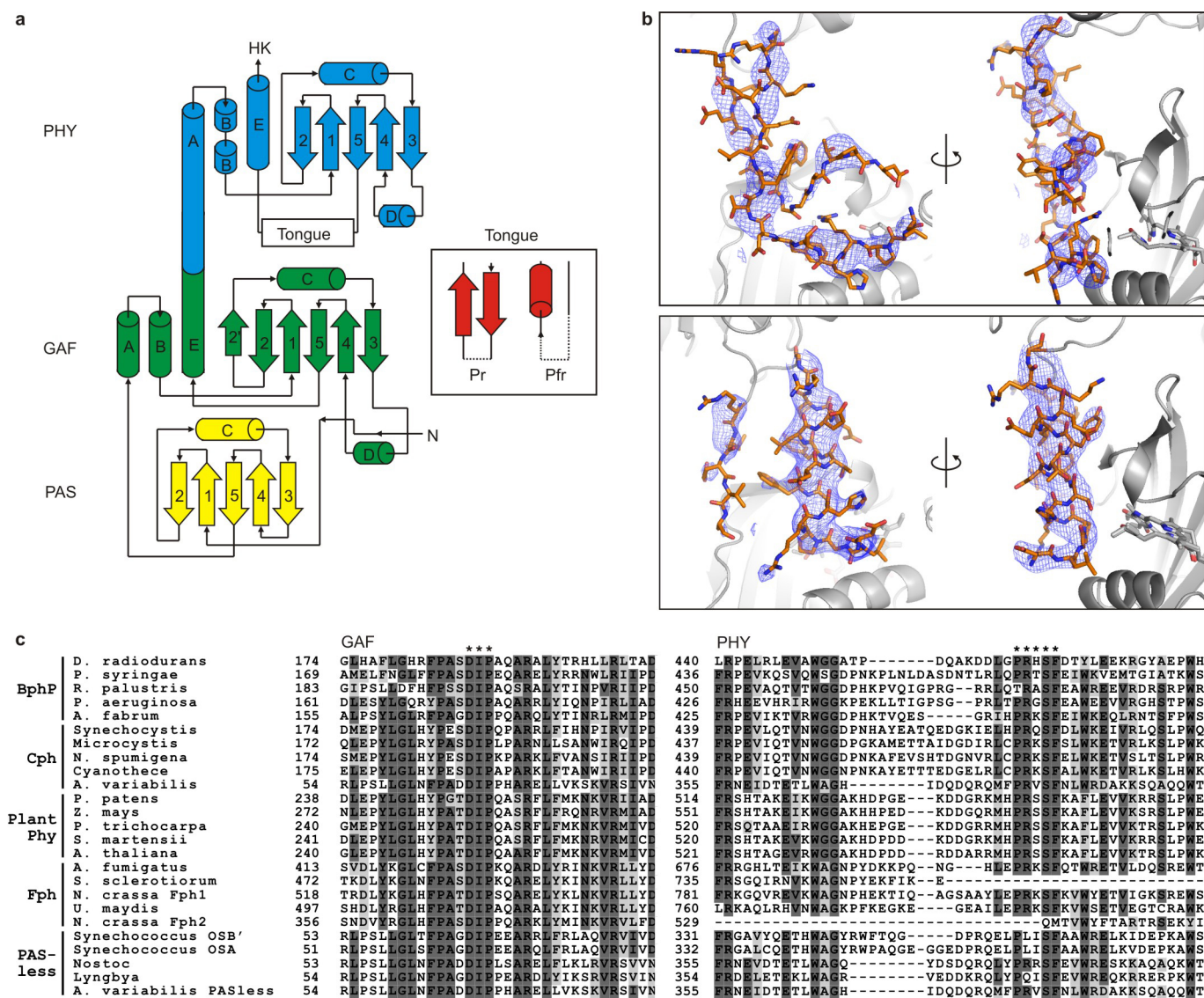
Author Information Coordinates and structure factor amplitudes have been deposited in the RCSB Protein Data Bank (<http://www.rcsb.org/pdb/home/home.do>) under accession codes 4Q01 and 4Q0P. The solution structures are available as Source Data. Reprints and permissions information is available at www.nature.com/reprints. The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to J.A.I. (janne.ihalainen@jyu.fi) or S.W. (westenho@chem.gu.se).

31. Sievers, F. *et al.* Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. *Mol. Syst. Biol.* **7** (2011).
32. Wagner, J. R., Zhang, J. R., Brunzelle, J. S., Vierstra, R. D. & Forest, K. T. High resolution structure of *Deinococcus bacteriophytochrome* yields new insights into phytochrome architecture and evolution. *J. Biol. Chem.* **282**, 12298–12309 (2007).
33. Mailliet, J. *et al.* Spectroscopy and a high-resolution crystal structure of Tyr263 mutants of cyanobacterial phytochrome Cph1. *J. Mol. Biol.* **413**, 115–127 (2011).
34. Konarev, P. V., Volkov, V. V., Sokolova, A. V., Koch, M. H. J. & Svergun, D. I. *PRIMUS*: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* **36**, 1277–1282 (2003).
35. Anders, K., Daminelli-Widany, G., Mroginski, M. A., Von Stetten, D. & Essen, L. O. Structure of the cyanobacterial phytochrome 2 photosensor implies a tryptophan switch for phytochrome signaling. *J. Biol. Chem.* **288**, 35714–35725 (2013).
36. Gasteiger, E. *et al.* in *The Proteomics Protocols Handbook* (ed. Walker, J. M.) Ch. 52, 571–607 (Humana, 2005).



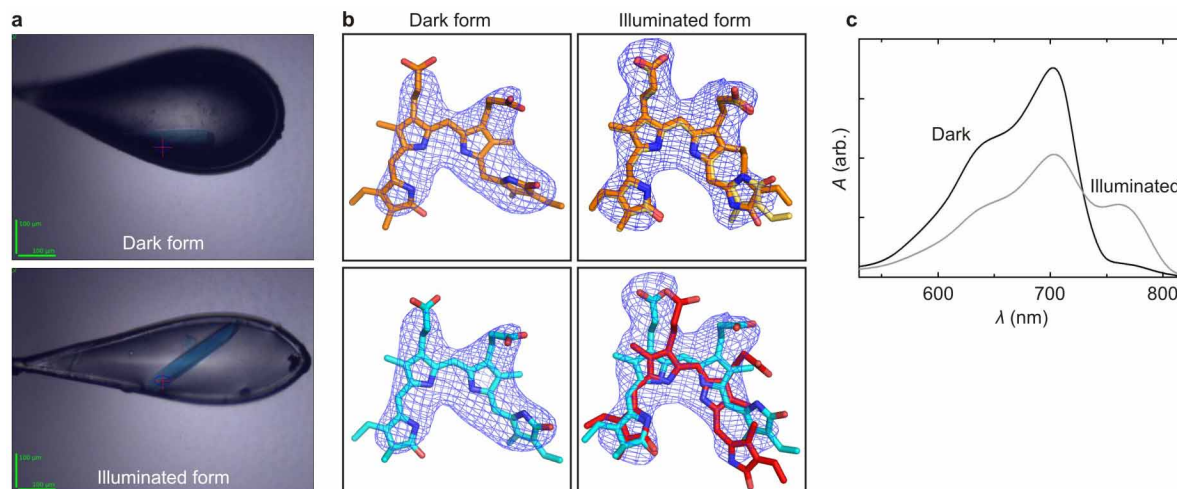
Extended Data Figure 1 | Difference scattering and kinetics. **a–c**, Singular value decomposition ($\Delta S(q, t) = USV^T$, where U , S and V are matrix factors) of time-resolved solution scattering data from the PAS-GAF-PHY fragment. Two components (n) suffice to describe the data, the final product ($n = 1$) and a transient low- q depression ($n = 2$). **a**, The first three basis spectra (first, second and third columns of US), and original X-ray scattering data (black) with reconstruction based on the first two singular values (all columns of USV^T , red) are shown. **b**, Relative amplitudes of the two first basis spectra (first and second columns of V). Lines are guides for the eye. **c**, Singular values (diagonal elements of S). **d**, The rise of the Pfr product state as measured by direct

integration of difference scattering curves ($\langle \Delta S(q)^2 \rangle_{1.2 \leq q \leq 2.5}$) and by optical transient absorption at 754 nm after excitation with a 680 nm laser flash. Solid curve is the function $(1 - e^{-t/\tau})$, where t is time and τ a free fitting parameter. These data establish that the structural change occurs just after the Pfr state is formed in the chromophore. Note the positive signal in the absorption curve with very small amplitude at $>3 \text{ ms}$, which appears to decay while the structural signal rises. This could be because the absorption properties depend weakly on the large-scale rearrangement. **e**, Direct static difference data from Fig. 1, amplified by q^2 to reveal wide-angle oscillations.



Extended Data Figure 2 | Light-induced changes in the secondary structure of the evolutionally conserved PHY tongue. **a**, Secondary structure and topology of the *D. radiodurans* PAS-GAF-PHY fragment. The structural elements in our crystal structures are very similar to other published phytochrome structures^{12,13}. The PHY tongue region (box), however, was found to refold upon illumination. The five-stranded β -sheet core of the GAF domain is extended by a small sixth β -strand (called 2') that interacts with the PHY tongue (see Fig. 2 and Extended Data Fig. 4). The mini-sheet structure at the knot region¹⁴ is not included in the graph. The figure is adapted from ref. 12.

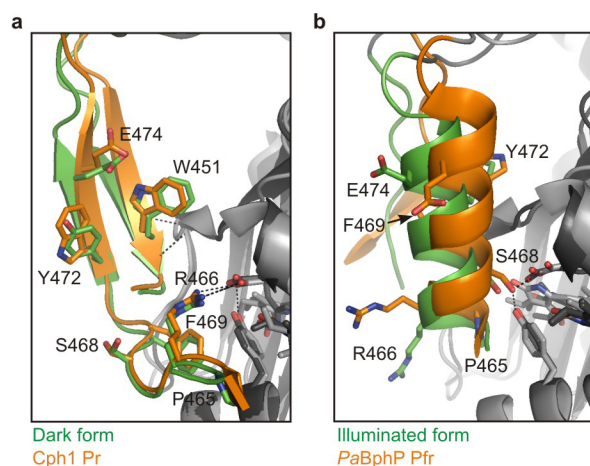
b, Omit map of the PHY tongue in the dark crystal form (top panel) and the illuminated crystal form (bottom panel). In the dark crystal form, the omit map density (blue) supports the built β -turn secondary structure (orange sticks), even though most of the side chains are poorly resolved. In the illuminated crystal form, the omit map (blue) clearly reveals the density of a helix with its bulky side chains (orange). The omit maps were calculated by repeating molecular replacement and a refinement step (see Supplementary Information) with a structure in which the PHY tongue was removed. All electron density maps are contoured at 3.0σ . **c**, Sequence alignment of part of the GAF domain and of PHY loop region. The conserved ²⁰⁷D-I-P²⁰⁹ motif in the GAF domain and ⁴⁶⁵P-R-X-S-F⁴⁶⁹ motif in the PHY tongue are marked by asterisks. Five representatives from eubacterial (BphP), cyanobacterial (Cph), higher plant, fungi (Fph) and PAS-less phytochromes are shown. The full alignment is supplied as source data. Colour coding: dark grey, identical residues; light grey, similar residues. The amino acid sequences were aligned with Clustal Omega using the default settings³¹. Sequences include *Deinococcus radiodurans* BphP, *Pseudomonas syringae* pv. tomato T1 BphP, *Rhodospseudomonas palustris* TIE-1 BphP3, *Pseudomonas aeruginosa* PAO1 BphP, *Agrobacterium fabrum* str. C58 BphP1 (Agp1), *Synechocystis* sp. PCC6803 Syn-Cph1, *Microcystis aeruginosa* NIES-843, *Nodularia spumigena* CCY9414, *Cyanobacteria* sp. PCC 7822, *Anabaena variabilis* ATCC 29413, *Physcomitrella patens* Phy1, *Zea mays* PhyB1, *Populus trichocarpa* PhyA, *Selaginella martensii* Phy1, *Arabidopsis thaliana* PhyA, *Synechococcus OSB'* SyB-Cph1, *Synechococcus OSA* SyA-Cph1, *Nostoc punctiforme* PCC73102, *Lyngbya* sp. PCC 8106 and *Anabaena variabilis* ATCC 29413.



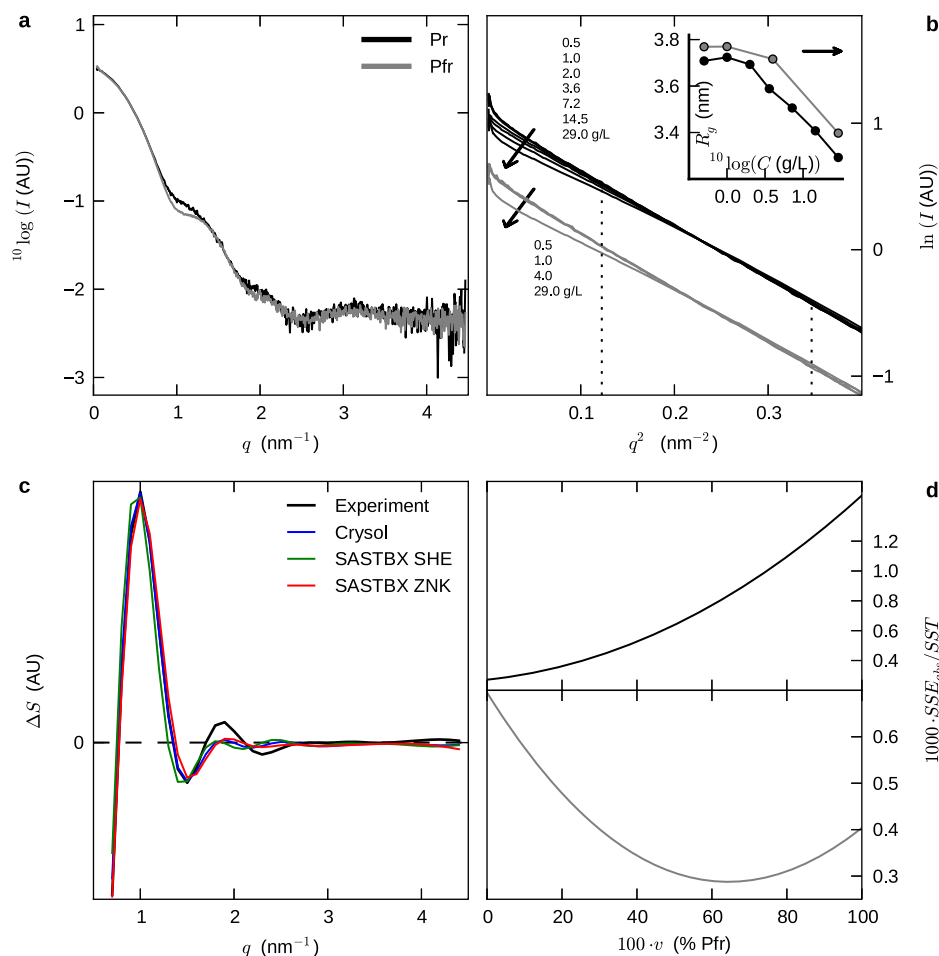
Extended Data Figure 3 | Biliverdin structure and spectra in crystals.

a, Photographs of the crystals under cryogenic conditions at the beamline ID23-1. **b**, Biliverdin omit maps of the dark (top left) and illuminated (top right) form support the existence of the modelled biliverdin conformations (yellow and orange). Comparison of the electron density around the biliverdin with published structures (bottom panels). In the dark form, the electron density indicates a conformation similar to the published Pr structures^{13,14,32,33}, including a *D. radiodurans* structure (PDB accession 2O9C, cyan)³². Therefore we modelled the biliverdin as a 15Za isomer. In the illuminated form, however, the electron density supports neither the biliverdin as determined in the Pfr structure of *PaBphP* (PDB 3NHQ, red)²⁰, nor as determined in the Pr structure of *D. radiodurans* (2O9C, cyan)³². Therefore the rotation of the biliverdin D-ring cannot be reliably determined and is modelled with both possibilities (15Za, and 15Ea, orange and yellow in top right panel). Omit maps were calculated as in Extended Data Fig. 2 and contoured at 3.0σ . **c**, Representative absorption spectra of the dark (black) and illuminated (grey) crystals, recorded at 123 K. Note that the terms 'illuminated' and 'dark' refer here to the crystallization conditions (see Supplementary Information for details). The dark crystal spectrum resembles the Pr spectrum in solution (Fig. 1a). Illumination with red light in the crystallization drops at ambient temperature

led to a slight increase of far-red absorption and disintegration of the crystals (data not shown). Dark crystals were unaffected by far-red illumination. The spectrum of the illuminated crystals shows that a substantial proportion (>50%) of the proteins reside in Pfr state. A similar absorption was detected from solubilized crystals. The illuminated crystals could be switched to Pr-like absorption with far-red illumination. Reversely, the Pfr-like features could also be increased with red light (data not shown) with illumination at ambient temperature. Exposure with light increased the scattering background in the absorption measurements. The crystals seemed unaffected by the illumination when illuminated with red light in the crystallization drops. Although the spectral analyses of the illuminated crystals do not indicate a pure Pfr spectrum, and the biliverdin conformation cannot be fit unambiguously to the electron density, the remainder of the electron density is homogeneous (Extended Data Fig. 2b). Most importantly, the tongue region of the PHY domain adopts the conformation resembling the Pfr state of *PaBphP* (Extended Data Fig. 4b)^{12,19,20}. The conformations of the four monomers in an asymmetric unit are practically identical and hence we conclude that biliverdin can co-exist in both Pr and Pfr states inside this crystal form and still the protein part represents the structural aspects of the Pfr state only.

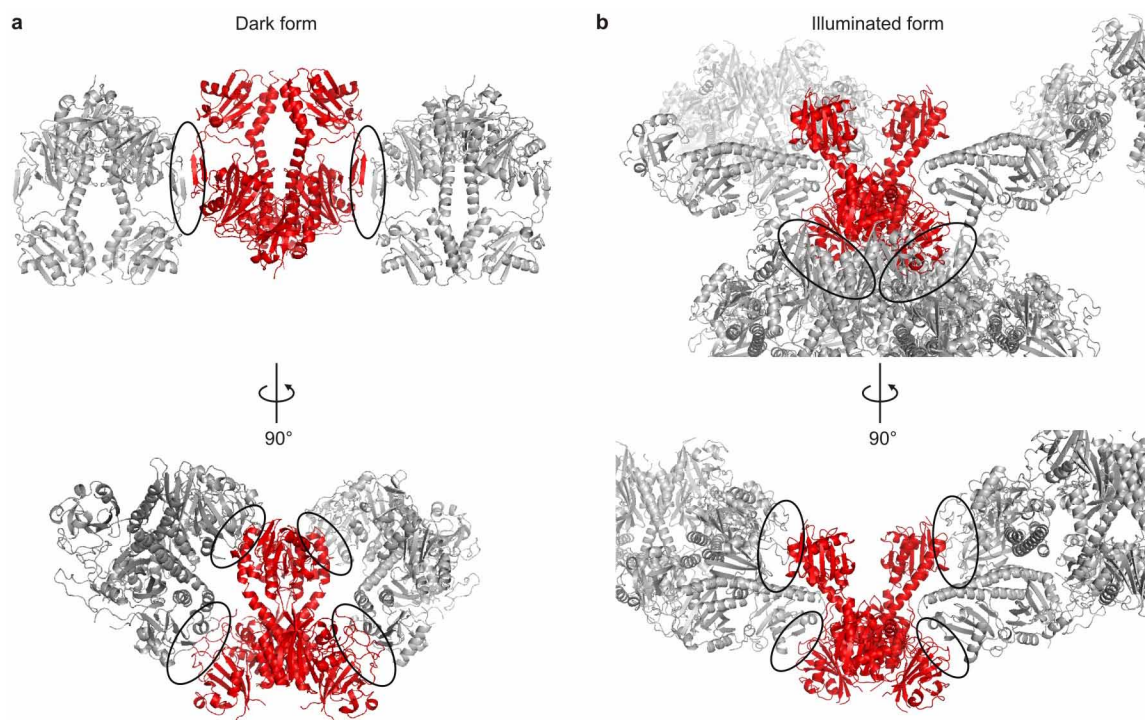


Extended Data Figure 4 | Comparison of PHY tongue interactions with published structures. **a**, Comparison of the dark crystal form (green/dark grey) to cyanobacterial *cph1* in Pr state (PDB 2VEA, orange/light grey)¹³. **b**, Comparison of the illuminated crystal form (green/dark grey) to PaBphP in Pfr state (PDB code 3NHQ²⁰, orange/light grey). In both the Pr and Pfr forms, key interactions are conserved between the phytochromes (black dashes), as well as the positions of three conserved tongue motives (see Extended Data Fig. 2c). The residues of these three motives are indicated as: ⁴⁵¹W-(G/A)-G⁴⁵³, ⁴⁶⁵P-R-X-S-F⁴⁶⁹ and ⁴⁷²(W/F,Y)-X-E⁴⁷⁴,¹³ with numbering from the *D. radiodurans* sequence. Trp 451 was not modelled in our illuminated crystal structure, and part of the PHY tongue has been removed for clarity. Small changes in relative orientations between the difference crystal structures are observed, for example, a slight tilt of helix of the Pfr tongue.



Extended Data Figure 5 | SAXS data and calculations. **a**, Experimental SAXS data of dark ('Pr') and pre-illuminated ('Pfr') samples. Here, I is scattered intensity, 2θ is the scattering angle and q is the magnitude of the scattering vector. The data are merged from the concentration series (Extended Data Table 2b) and are normalized on $0.4 \text{ nm}^{-1} < q < 0.6 \text{ nm}^{-1}$. **b**, Guinier plot of the low- q region, shown for all concentrations. The curves converge at low concentration (C). Inset shows the radii of gyration (R_g) calculated from the curves in **a** according to the Guinier approximation. Dotted lines show the q -range used for linear regression ($0.15 \text{ nm}^{-2} < q^2 < 0.120 \text{ nm}^{-2}$), which is such that $qR_g < 1.3$ as required³⁴. Arrows indicate increasing concentration. **c**, Average difference scattering signals calculated from the solution-structural models using three methods: Crysol (default settings), SASTBX with spherical harmonic expansion (SHE, default settings), and SASTBX with Zernike

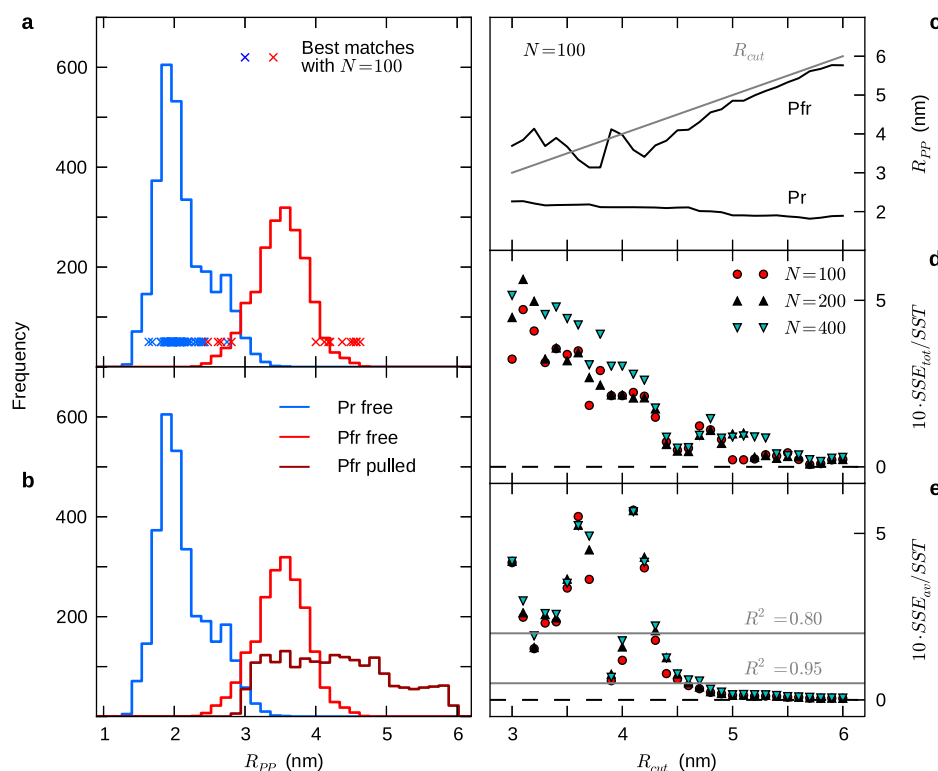
polynomial expansion as described in Supplementary Information³⁰. Different calculation methods result in very similar predicted difference X-ray scattering patterns. **d**, Determination of the relative Pr/Pfr populations represented by the BM29 data as described in Supplementary Information. SSE_{abs} is the sum-squared error against absolute scattering data, SST is the sum of squares about the mean, and v is the Pfr fraction. We find that our Pr sample contained only Pr (top) whereas in the Pfr sample 64% of the protein molecules adopted the Pfr conformation (bottom). Notably, the static difference scattering patterns (Fig. 1) represent, up to a scaling factor, the relation between pure Pr and Pfr populations. This is in contrast to traditional SAXS which report on population mixtures, because the Pfr state cannot be easily produced with 100% population in solution.



Extended Data Figure 6 | Packing interactions of the crystal forms.

a, b, Crystal packing interactions of the dark (**a**) and illuminated (**b**) crystal forms. The dimer of an asymmetric unit is shown in red and the symmetry mates in grey. For the illuminated form, only chains A and B are shown in red. Interactions are shown from two orientations for clarity. In the dark form, crystal contacts are seen in the top regions of the PHY domains and therefore may cause artefacts in the long scaffolding helix and in the opening of the PHY domains. In the illuminated form contacts are such that the PHY domains may be pushed closer together, which is consistent with the larger separation of the

PHY domains as refined from the solution X-ray scattering data. It is noteworthy that the relative orientation of the monomers in the dimer is different between all three known structures for PAS-GAF-PHY phytochromes. For *P. aeruginosa* the dimer is parallel with variations in between different copies of the dimer in the crystallographic unit cell¹², in two cyanobacterial phytochrome an antiparallel dimer is observed^{13,35}, and in our Pr structure, the monomers have an angle of approximately 45°. The rings highlight crystal contacts.



Extended Data Figure 7 | Solution-structural refinement. **a**, The distribution of PHY domain separations (R_{PP}) obtained from unbiased molecular dynamics simulations (production runs 1–3). The best-matching sets *A* and *B* of solution structures found from these conformations are indicated by crosses (with $N = 100$ best pairs chosen from $M \approx 6 \times 10^6$ possible combinations). Whereas the *Pr* structures cluster in a region of high sampling, the *Pfr* structures lie at the edges of the PHY-PHY distribution, suggesting inadequate sampling. To remedy this, we artificially scanned the PHY domain separation in separate simulations (production runs 4 and 5) to improve sampling. **b**, The new distribution of pulled PHY domain separations in the *Pfr* state. The final analysis and all solution-structural conclusions drawn in this study are based on the trajectories described in **b**. **c–e**, Consistency test of the structural refinement procedure. **c**, A cutoff parameter R_{cut} was introduced

to reject all molecular dynamics frames $R_{PP} > R_{cut}$. The resulting average over R_{PP} of the best $N = 100$ pairs is plotted as a function of R_{cut} . It is found that $R_{PP} \approx R_{cut}$, which indicates that the best fit to the difference X-ray scattering data are always at the highest separations available in sampling range. **d** and **e** show the dependence of the total and average error (SSE_{tot} and SSE_{av}) as functions of R_{cut} . SST represents the overall sum of square about the mean. It is observed that the error decreases steeply for $R_{cut} < 5$ nm, and only marginally for $R_{cut} \gtrsim 5$ nm. We therefore consider optimization in the latter range overfitting, and applied $R_{cut} = 5.0$ nm in the refinement for the solution structures. The refined structures (Fig. 4) should therefore be considered to represent lower limits on R_{PP} . **f**, Solution-structural refinement parameters and results.

Extended Data Table 1 | Crystallographic data collection and refinement statistics

	Dark form*	Illuminated form*
Data collection		
Space group	P6 ₄	P2 ₁ 2 ₁ 2 ₁
Cell dimensions		
<i>a</i> , <i>b</i> , <i>c</i> (Å)	150.28, 150.28, 145.95	86.90, 195.70, 225.00
α , β , γ (°)	90.0, 90.0, 120.0	90.0, 90.0, 90.0
Resolution (Å)	48.65-3.80 (3.90-3.80) [†]	49.22-3.24 (3.32-3.24) [†]
<i>R</i> _{merge} (%)	39.2 (451.3)	24.1 (439.0)
<i>I</i> / σ	4.72 (1.05)	9.69 (0.83)
Completeness (%)	99.9 (100.0)	99.9 (100.0)
Redundancy	11.0 (11.6)	19.3 (20.3)
Refinement		
Resolution (Å)	48.65-3.80 (3.90-3.80)	49.22-3.24 (3.32-3.24)
No. reflections	17597 (1295)	58759 (4255)
<i>R</i> _{work} / <i>R</i> _{free}	25.3/26.6 (37.10/40.20)	23.3/26.0 (39.7/39.7)
No. atoms		
Protein	7416	14784
Ligand/ion	86	172
Water	0	0
Average B-factor (Å ²)	210.24	158.61
Protein	208.71	156.92
Ligand/ion	170.21	143.48
Water	-	-
Wilson B-factor (Å ²)	130.97	118.65
R.m.s deviations		
Bond lengths (Å)	0.006	0.005
Bond angles (°)	1.112	1.159

*Data from two crystals were used.

[†]Highest resolution shell is shown in parenthesis.

Extended Data Table 2 | SAXS statistics and sample details

A

	PAS-GAF-PHY		Bovine serum albumin (BSA)
	Dark	Illuminated	
R_g (nm)	3.75 ± 0.02	3.78 ± 0.02	3.10 ± 0.01
R_g (nm)*	3.72	3.77	
D_{\max} (nm)	11.5	12.4	
V_{Porod} (nm ³)	173	165	
I_0	97.3 ± 0.2	86.8 ± 0.3	64.3 ± 0.03
MW_{exp} (kDa)	~100	~90	~66
MW_{calc} (kDa)	113.0	113.0	

* from manual analysis in Extended Data Figure 5

B

Sample	Conc. [mg/ml]	Buffer	Salt	pH
cSAXS				
PAS-GAF	18	20 mM Tris	75 mM NaCl	7.8
PAS-GAF-PHY	25	20 mM Tris	150 mM NaCl	7.0
BioCARS				
PAS-GAF	28	30 mM Tris		8.0
PAS-GAF-PHY	25	30 mM Tris		8.0
BM29				
PAS-GAF	0.5-26*	30 mM Tris		8.0
PAS-GAF-PHY	0.5-29*	20 mM Tris	150 mM NaCl	7.0

*see Extended Data Figure 5

a. Statistics of the static SAXS (BM29) data, including radii of gyration (R_g), maximum particle dimension (D_{\max}), Porod volume (V_{Porod}), forward scattering (I_0). The molecular weights (MW_{exp}) were estimated from the I_0 of BSA using the formula $MW_{\text{exp}} \approx MW(\text{BSA}) * [I_0(\text{sample})/I_0(\text{BSA})]$, where $MW(\text{BSA}) = 66$ kDa. Theoretical molecular weights (MW_{calc}) were calculated from the protein sequence³⁶. **b.** Details of the samples used for solution scattering measurements.

CAREERS

TURNING POINT Neuroscientist reaches milestone in fruit-fly brain map **p.251**

WORK-LIFE BALANCE Creativity breaks can pay job dividends, study finds **p.251**

NATUREJOBS For the latest career listings and advice www.naturejobs.com

FANATIC STUDIO/GETTY



CONTINUING EDUCATION

Study broad and deep

Short tutorials can help researchers to sharpen skills, and longer courses can provide a way to deepen knowledge.

BY SARAH WEBB

A successful research career demands that scientists engage in non-stop learning — after all, technologies advance, interests shift and discoveries transform understanding, especially in today's era of 'big data'. Training sessions such as workshops, courses or online tutorials can fill the gaps in a scientist's knowledge and skills, helping them to improve and expand their research programme and support their efforts to land a job.

To derive maximum benefit from off-site training, researchers need to identify exactly

where the gap is, then search online or talk to colleagues to find a workshop or course that matches their needs — scientific societies and conferences are often good sources — and manage logistics, including cost and timing (see 'The right fit'). They may want to apply for grants or fellowships from societies or other professional organizations to cover tuition and off-site travel expenses, and allocate the time needed to do so. They may have to calculate how to juggle their resources if they are unable to get funding or if their employer cannot provide support. Some courses are offered only once a year and may fill up quickly, so prospective students need to make themselves aware of

enrolment dates and be ready to pull the trigger.

When does a short course or workshop make sense? Researchers often turn to training manuals or online tutorials when they need to pick up a new laboratory technique or learn how to use equipment or software. But sometimes hands-on, in-person instruction is more effective than reading, especially when it comes to mastering complex tasks such as troubleshooting instrument problems or managing data analysis.

Are the costs worth it? Tuition might cost anything from a few hundred dollars for a day-long workshop to US\$1,500 for a week-long course — and that would not include travel ►

► expenses. (Some day-long workshops and courses are scheduled directly before or after major scientific meetings, allowing researchers to save on travel costs.)

Furthermore, if the researcher is lucky, the right offering will be close by. Saik-Kia Goh, a bioengineering PhD student at the University of Pittsburgh in Pennsylvania, took a free two-week course in clinical and translational research at the US National Institutes of Health (NIH) in Bethesda, Maryland, an easily drivable distance of 370 kilometres. Accommodation costs came to about \$2,000, half of which he covered himself — but he not only gained a greater understanding of the clinical and regulatory side of stem-cell research, he was also able to discuss his research with NIH scientists and learn about postdoctoral opportunities he had not been aware of. “I wasn’t expecting it to be as much of a networking opportunity,” he says. “It was an investment in my career.”

Sometimes, however, researchers need to travel halfway around the world. Kirstin Wurms, a plant pathologist at Plant & Food Research, a government institute in Hamilton, New Zealand, examines mechanisms of plant resistance to disease. The field increasingly relies on gene-expression data from a technique known as the quantitative polymerase chain reaction (qPCR), which requires specialized expertise to produce reliable, reproducible data. Wurms, who did not have that know-how, initially relied on a colleague’s assistance. But when that colleague left the institute, she knew that she needed to learn the technique herself. She signed up for two short courses offered last September at TATAA Biocenter, the world’s largest qPCR training centre and service provider, in Gothenburg, Sweden. Supported by a fellowship that helps to pay for overseas travel for scientists, plus some funds from her institute, Wurms learned the technique, its international guidelines and how to use a statistical program to analyse data.

International travel was also on the cards for Stefan Suter, a wildlife-management researcher at the Zurich University of Applied Sciences in Switzerland. He needed to learn how to analyse communication between wild boars, and found that the Cornell Lab of Ornithology in Ithaca, New York, offered a week-long course on working with sound-analysis software. Suter had used a simplified version of the software during his doctoral research, but needed to go deeper to realize his plan of protecting



“It was an investment in my career.”

Saik-Kia Goh

Trying to figure out how to fill a knowledge gap can be overwhelming. Jacquelyn Gill, a palaeoecologist at the University of Maine in Orono, has advice on how to tackle the issue.

- Look around online and talk to colleagues and mentors about the skills you would like to learn and how you might bridge the gap.
- Know your learning style. If you work well independently, you might be able to learn about a statistical model or a programming language simply by working through a textbook. If you need group accountability, you might be able to work with a group of other graduate students or postdocs who

want to learn the same skill.

- If money is tight or travel difficult, consider an online option.
- Sometimes, an in-person course is the only way to gain a particular skill. If that skill is needed at a crucial point in your training, you should make a good case for the time and expense.
- “When you apply for funds, you need to get really good at framing why that course is essential for you at this stage in your career,” Gill says. If you can’t communicate that need effectively, maybe you do not need the course as much as you thought. **S.W.**

agricultural fields. The course helped him to create warning calls that scare boars away from valuable crops. Suter received full financial support from his university for the US trip because he made the case that taking the course would also allow him to improve his English. His previous experience working with the software meant that Suter could focus on specific questions and maximize the opportunity to get answers from experts. Although the course had a basic curriculum, “there was a lot of space for individual questions, and an interactive approach, which I really appreciated”, he says.

UNEXPECTED BENEFITS

Short courses can be effective for quickly acquiring new skills. But at other times — when aiming to deepen their knowledge of a research area or to explore a new field — a researcher may want to devote more than a week or two to their pursuit, even though the costs of tuition, travel and accommodation are likely to be much higher. Four to eight weeks might sound like a long time to be away from one’s primary research, but it can provide an opportunity to dig more deeply into a discipline or to expand into a new field.

Physiologist Christopher Tubbs used a six-week seminar at the Marine Biological Laboratory in Woods Hole, Massachusetts, to deepen his knowledge of reproductive biology. Tubbs, who now examines the effects of endocrine disruption on endangered wildlife, took the seminar as a doctoral student in the hope of learning new ways to study fish-sperm motility.

At the same time that he learned more about fish, he acquired a new understanding of hormones, hormone signalling and gamete biology that helped him to secure his current post at the San Diego Zoo Institute for Conservation Research in California.

Savvy researchers can often find ways to bring down costs to achievable levels. Chemist

Jennie Mayer of Bellevue College in Washington wanted to learn how to maintain a new mass spectrometer and nuclear magnetic resonance spectrometer purchased by her institute. Fortunately, Mayer subscribed to e-mail updates from the American Chemical Society (ACS), of which she was a member, and received one about its six-week series of online chemistry courses known as Sci-Mind. These include tutorial videos, laboratory exercises and opportunities to interact with instructors and other students.

At its original price, the course would have cost just half that of the on-site equivalent, and she managed to get both the members’ rate and a promotional discount that ACS was offering, with a final price tag to her institution of about \$900. The course refreshed her skills on how to use the instruments and taught her how to maintain and troubleshoot the spectrometers. She is now creating modules that will help her to train students and her colleagues in their use.

As a postdoc, palaeoecologist Jacquelyn Gill opted to take a massive online open course (MOOC) in calculus because it was free and she could fit it into her schedule as she pleased. She had come to realize that she would need much deeper familiarity with calculus for her field’s increasingly quantitatively based research, but enrolling in a class alongside undergraduates felt awkward. So when the MOOC format became available, she leapt on-board and signed up for a 14-week course through the online platform Coursera. Each week, she spent just two hours working on problems and taking online quizzes.

Now a faculty member at the University of Maine in Orono, Gill stresses that it is never too late to learn something new — and far too easy to put it off. “Just do it,” she says, “and do it now.” ■

Sarah Webb is a freelance writer based in Chattanooga, Tennessee.

TURNING POINT

Marta Zlatic

Marta Zlatic, a neurobiologist at the Howard Hughes Medical Institute's Janelia Farm research campus in Ashburn, Virginia, has published an 'atlas' of the larval fruit-fly brain that pinpoints the neurons responsible for recorded behaviours. It is a major step towards realizing her dream of understanding how brains mediate behaviour.

What were your early interests?

Growing up in Croatia, I had fantastic Latin and Greek teachers. I loved grammar and the rules of language, learned eight languages and enjoyed reading literature in its language of origin. But I was also interested in biology, stemming largely from wanting to understand the neural basis of language and behaviour. I studied linguistics and Russian at the University of Zagreb in Croatia before starting at the University of Cambridge, UK, where I was able to explore my interests in neuroscience, languages and theatre.

What were your pivotal career moments?

There were a couple. One was attending lectures on the assembly of neural circuits in fruit-fly embryos by Cambridge neuroscientist Michael Bate, who inspired me to study the first steps of that process. I ended up doing both my PhD and postdoc at Cambridge. Another was receiving a prestigious postdoctoral fellowship from Cambridge's Trinity College, which allowed me to freely pursue research in neural-circuit assembly at both Cambridge and abroad. I spent a year at Columbia University in New York and a year at Janelia Farm, where I discovered that colleagues were developing amazing genetic tools, including transgenic fruit flies in which you could manipulate neuron types, which is necessary to create a functional neuroanatomy map.

How did working at Janelia affect your career plans?

My husband, Albert Cardona, and I met as visiting postdocs at Janelia. I was working out how to piece together a neural-behaviour map, and he was determining the viability of a connectome, a map of neural activity in the larval fruit-fly brain. He got a job in Zurich, and I was hired as a group leader at Janelia. We came up with long-distance strategies such as working as visiting scientists at each other's institutes until Albert was hired as a group leader in neurobiology at Janelia in 2011. Our aim is to put the neural behavioural map and the connectome together



— an important step towards linking behaviour to the underlying brain circuitry.

Why start with fruit-fly larvae?

Fruit-fly larvae have only about 10,000 neurons, compared with the adult's 100,000, so it is a manageable system. Plus, we could experimentally control most neuron types of the larval neurosystem using existing genetic tools.

How did you put the atlas together?

Janelia was one of the few places in the world where this research could have been developed. We had access not only to transgenic fruit flies, but also to images of neuronal expression patterns in those larval brains. In collaboration with Janelia colleagues, we developed a system that could generate movies of larval behaviour in 40,000 animals in response to stimulation of specific neurons. We also worked with Carey Priebe at Johns Hopkins University in Baltimore, Maryland, who created a program that could distinguish between 29 types of behaviour and map each to specific neurons. Altogether, creating the atlas took three years of collecting data, making sense of them and verifying them.

How will the atlas affect your career?

This is the foundation for the work I want to continue with Albert. We need to combine three maps: of neuron-behaviour, neuron connectivity and neuron activity. Together, they will lay the groundwork for a brain-wide understanding of behaviour. If we can one day understand how complex sequences of actions are generated and learned in model organisms, maybe we could look for similar principles in humans. ■

INTERVIEW BY VIRGINIA GEWIN

WORK-LIFE BALANCE

Learn from leisure

All work and no play could make for a mediocre researcher, according to a study (K. J. Eschleman *et al.* *J. Occup. Organ. Psychol.* <http://doi.org/sfg;2014>). Workers who also engage in creative activities such as painting or playing music perform more effectively in their job, says lead author Kevin Eschleman, a psychologist at San Francisco State University in California. Creative activities outside work also help employees to bounce back from stress and to learn new skills, he says. Eschleman and his co-authors surveyed 433 US workers in various occupations, including 92 active-duty US Air Force captains. Eschleman advises scientists to take up leisure activities to help their conscious mind to detach from workplace pressures and their subconscious to find different approaches to their research.

UNIONS

Adjunct action

A concerted effort by an international union to link up adjunct US faculty members is succeeding in its city-by-city approach. The Service Employees International Union has enlisted about 70% of adjunct faculty members in the Washington DC area. Malini Cadambi, the union's higher-education campaign director, says that 80% sign-up is possible. The union last year launched similar strategies in Washington state and Massachusetts, and is expanding to seven more US regions. A wide membership base provides adjunct faculty members with more bargaining power in their push to raise pay, make evaluations fairer and boost academic freedom, says Cadambi.

REMUNERATION

Narrowing pay gaps

Gaps between vice-chancellors' and professors' pay rises reported in the *Times Higher Education* 2012–13 salary survey were highlighted as UK faculty members negotiated contracts for the upcoming academic year. The survey found that professors' salaries rose by an average of just 1%, to £75,472 (US\$126,782), whereas vice-chancellors' pay (including some benefits) jumped by 5.5% to £232,120. This year, professors have been offered a 2% rise, says Jenny Sherrard, a spokeswoman for the University and College Union in London. "It's not acceptable for those at the top to be receiving higher percentage pay rises than are on offer to the rest," she says.

KILLSTEALER

A shot in the dark.

BY ALTER S. REISS

“Will you put that game down!” snapped Steve, and Brendan turned to his head, looking surprised and hurt.

Steve was immediately sorry. He couldn't yell at what was happening, but snapping at the kid it was happening to wasn't fair. Brendan was just 12, old enough to understand what was going on, but not old enough to ... not old enough for anything. He looked like he hadn't slept; bruises under his eyes, skin like paper. He looked like he was dying.

“Sorry,” said Steve. “Just ... stay with me for a little bit, okay?”

“Okay,” said Brendan. “But let me wrap things up. My friends will —” and he looked back at his game and started making excuses into his headset.

Friends, Steve thought. *Like the kid had any real friends — just people on the net. I should've pushed harder, with school, with summer camp, things like that. Too late now.* He breathed, tried to focus. His job was to be Brendan's father, for however much longer Brendan had.

“There. What is it, Dad? What're you printing?”

There was the click and whirr as the printer finished its job. Best system on the market — laser sintering for four metals, 14 elements total for non-metallic compositions. Should have spent his time with his kid instead of with his toys. “Bots,” he said. He took out the capsule it had printed, and passed it over to Brendan. “Here, take this.”

Brendan frowned, but took it and swallowed it down with a cup of water. He'd been given a lot of pills to take since they'd found the problem, but he was never a complainer.

By the time Steve had the software set up, the bots were already in Brendan's bloodstream, and the boy was leaning on his shoulder, interested. “Thought it might help to see what's happening,” said Steve, which wasn't exactly true. The book he'd been reading recommended showing kids what was going on, and he thought that sounded reasonable.

The screen revealed a confusion of shapes, all moving at different speeds. “That's your blood, you see,” he said.

“And those are the cancer cells?”



Steve winced, wishing Brendan hadn't used that word. “No,” he said. “The ones outlined in red are the red blood cells. There, that one in yellow is a white blood cell. Some of them are the ... some of them aren't working right.”

“Like that one?”

“Yeah,” Steve said. “The hairy-looking ones are the bad cells. And check this out.” He pressed some buttons, and the bot went forward, into the cell, and it fell apart.

“Cool,” Brendan slid in next to him, took over the keyboard.

“If we could do chemo, the bone marrow would stop making —”

“But we can't,” said Brendan. “Because of my liver. And we tried other things that didn't work. So I'll pop them.”

Steve shook his head, trying not to cry. It had seemed reasonable in the book, but it wasn't going to be easy. “There are too many of them to get. Tens of millions. But you can get a few, give them a fight.”

“Thanks, Dad,” said Brendan, settling in more comfortably. “Maybe I'll get some of my friends to help — are these things networked?”

“I'm not sure,” said Steve. He wanted to

stay. There wasn't anything he could do, but he wanted to stay. Only, while their insurance was rubbish, if he lost his job, it'd be gone too. “Look, I have to get to work; if you find a better pattern, you can print it out.”

“Thanks,” said Brendan, but already he wasn't listening. Trade one game for another, but the book Steve had been reading recommended it as a first step, and it seemed to get Brendan interested.

When Steve got back home, Brendan was still popping cancerous lymphocytes. Maybe it was a mistake, letting him think that he'd be able to beat leukaemia one cell at a time. The idea in the book was that Brendan would get frustrated and stop, and get angry and come back to it, and eventually get used to the idea of how big the problem was.

“How's it going?” Steve asked.

“Oh, that's not me,” said Brendan, pulling back, and giving his father a half smile. “Once they started getting bonus points for the spawn sites —”

“The bone marrow?”

“Yeah, the screwed up bone marrow. Even with the cascading bonus for the last few, you can't really get that many points working on me. There're still some guys going for those bonuses, but I'm working on some guy in Fiji who just started printing.”

Steve tried to process. “There are tens of millions of cancer cells in your blood,” he said.

“Were,” said Brendan. “My friends got the interface and the targeting better, and their friends wanted to help too. There's a leaderboard and there are a lot of people playing.”

“You're better?” asked Steve, turning Brendan around for a closer look. He still looked like he hadn't slept, but his eyes were clear, and maybe his skin was even a little better.

“Yeah, but this guy ... Dad, you're crushing me. Dad! Oh, man, someone's stealing my kills, Dad!” ■

Alter S. Reiss is a field archaeologist and scientific editor. His fiction has appeared in Fantasy & Science Fiction, Strange Horizons, Daily Science Fiction and elsewhere.

➔ **NATURE.COM**
Follow Futures:
@NatureFutures
go.nature.com/mtoodm